

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:03:45 ; Search time 32.7914 Seconds
(without alignments)
10995.308 Million cell updates/sec

Title: US-09-833-740-7_COPY_1_210
Perfect score: 210
Sequence: 1 ccgcctgtctcttcctcc.....ggattgtgcaacattcc 210

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.2	29.6	2170	AAH43758	5' flanking and 5'
2	38.6	18.4	772	AAH04015	Human cDNA clone (
3	38.6	18.4	1612	AAH15417	Human cDNA sequenc
4	34.2	16.3	13327	AAH02939	Human reproductive
5	33.4	15.9	496	AAH07185	Human reproductive
6	33.2	15.8	1302	AAH60795	Human polynucleoti
7	33.2	15.8	1732	AAH59009	Human polynucleoti
8	33.2	15.8	1601	AAH83074	Human transcriptio
9	33.2	15.8	1831	AAH94639	Human full-length

10	33.2	15.8	7601	AAH36532	Human musculoskele
11	33.2	15.8	7601	AAH36533	Human musculoskele
12	33.2	15.8	11360	AAH03820	Human reproductive
13	33.2	15.8	11360	AAH04727	Human reproductive
14	32.4	15.4	553	AAH94801	Drosophila melanog
15	32.4	15.4	555	AAH92773	Human cDNA 3'-end
16	32.4	15.4	1194	AAH41329	Human normal uteru
17	32.4	15.4	1194	AAH72250	Drosophila odorant
18	32.4	15.4	1276	AAH05803	Drosophila melanog
19	32.4	15.4	1285	AAH94846	Drosophila melanog
20	32.4	15.4	3346	AAH05802	Human full-length
21	32.4	15.4	3464	AAH94338	Drosophila melanog
22	32.4	15.4	4345	AAH05370	Drosophila melanog
23	32.4	15.4	7632	AAH21136	Human reproductive
24	32.2	15.3	9370	AAH07075	TR12 related DNA-1
25	31	14.8	35414	AAH00147	cDNA encoding nove
26	30.8	14.7	555	AAH57072	Human immune/haema
27	30.8	14.7	555	AAH56984	DNA encoding novel
28	30.8	14.7	4264	AAH27635	DNA encoding novel
29	30.8	14.7	4264	AAH27636	Human immune/haema
30	30.8	14.7	4264	AAH68503	Human immune/haema
31	30.8	14.7	4264	AAH68504	Human immune/haema
32	30.8	14.7	4264	AAH69247	Human immune/haema
33	30.8	14.7	4264	AAH69248	Human immune/haema
34	30.8	14.7	16595	AAH33521	Human kidney amino
35	30.6	14.6	4136	AAH60231	Stef/Gbeta intera
36	30.6	14.6	10529	AAH09028	Maize fluoxyz gene
37	30.4	14.5	977	AAH48603	Arabidopsis thalia
38	30.4	14.5	980	AAH35669	Human reproductive
39	30.2	14.4	2073	AAH04589	Human nervous syst
40	30	14.3	11711	AAH18962	Human nervous syst
41	30	14.3	11711	AAH18963	Human nervous syst
42	29.8	14.2	394	AAH08267	Human mnti collage
43	29.8	14.2	670	AAH87950	EST clone FH136.
44	29.8	14.2	7958	AAH234835	Mouse integrin sub
45	29.8	14.2	16489	AAH27823	DNA encoding novel

ALIGNMENTS

RESULT 1
AAH43758
ID AAH43758 standard: DNA; 2170 BP.

AAH43758;

30-JAN-2002 (first entry)

5' flanking and 5' UTR of GLP-2 receptor gene.

5' flanking region; 5' UTR; untranslated region; murine;
glucagon-like peptide-2; GLP-2; receptor; regulator;
intestinal epithelium; lateral hypothalamus; promoter; ds.

Mus musculus.

Location/Qualifiers
Key
misc_RNA
misc_RNA
CDS
W0200179290-A2.
25-OCT-2001.
/tag= a
/note= "Punitive transcriptional start site"
/tag= b
/note= "Corresponds to translational start in rat/human
GLP-2R gene"
/tag= c
/note= "Putative translational start site"

PF 12-APR-2001; 2001WO-IB00619.
XX
PR 13-APR-2000; 2000US-196909P.
PR 01-FEB-2001; 2001US-265310P.
XX
PA (ONEO-) 1149336 ONTARIO INC.
XX
PI Drucker DJ, Lovshin JAL;
XX
WI: 2002-026012/03.
DR
XX
PT Novel recombinant DNA construct, useful as drug screening tool and as
PT protein delivery vehicle, comprising glucagon-like peptide-2 receptor
PT gene, and linked to heterologous gene of interest for expression
XX
XX Disclosure; Fig 1; 70pp; English.
XX
CC This sequence represents the 5' flanking and 5' UTR of the murine
CC glucagon-like peptide-2 (GLP-2) receptor gene. This sequence may be
CC used in the DNA construct of the invention, such that it is linked
CC for expression with a heterologous gene of interest. This construct
CC is useful for screening compounds to identify regulators of GLP-2
CC receptor expression. The DNA construct is also useful for delivering
CC a gene product to tissues selected from intestinal epithelium and the
CC lateral hypothalamus, by transfecting an organism, or a gametic or
CC embryonic form of such an organism, with the construct. The DNA
CC construct is also useful to identify cells capable of mediating
CC expression from the chosen GLP-2R promoter, to identify regions of
CC the GLP-2R promoter that are functional in a given cell type, and
CC to screen for agents that modulate expression from the GLP-2R promoter.
CC It can be used to map functional regions of the GLP-2R promoter.
CC for screening agents that modulate the function of GLP-2R promoter.
XX
SQ Sequence 2170 BP; 574 A; 501 C; 550 G; 540 T; 5 other;
Query Match 29.6%; Score 62.2; DB 24; Length 2170;
Best Local Similarity 67.5%; Pred. No. 1.3e-09;
Matches 141; Conservative 0; Mismatches 48; Indels 20; Gaps 3;
QY 3 GCCTTGTCTTCTCTCCTCAGCTGTCAAGAGTCCGAGAACACACAGTACTT-AGGG 61
DB 1476 GTCTTCTCTTCTCTCTGCTGTGAGAGAGTCCAGCAGCAGTACGTCTTGGGG 135
QY 62 AAGCTGTGGGAAAAATCTCCCTGCTTTGGGGGGGCGAGGGGCGGATGACCGAGGCC 121
DB 1536 TAGCTTGGGAAAAATCTCCCAAGATTCTTAGAGGGGCGAGCGGGGATGA----- 1586
QY 122 GAGAGGAACCTCTGAGACTCCGTAGATTGCTTAGACCGCCCTCAGACTCTCGGGCA 181
DB 1587 -----GAAACTGTGAGATTGCTGTAGATCGCTTAGAGCACTCAACAGTC-----GGCG 1636
QY 182 CGCTGAGAGGATTTTGTGCAACATTTC 210
DB 1637 GCCTGAAGAGGACTGTGCAACACTTCC 1665

RESULT 2
AAH04015/C
ID AAH04015 standard; CDNA; 772 BP.
XX
AC AAH04015;
XX
XX 26-JUN-2001 (first entry)
DT
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:850.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.

XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WI: 2001-318749/34.
DR
XX
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
PS Claim 1; SEQ ID 850; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 772 BP; 147 A; 258 C; 194 G; 170 T; 3 other;
Query Match 18.4%; Score 38.6; DB 22; Length 772;
Best Local Similarity 52.1%; Pred. No. 0.023;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 37 CCCAAGAACACAGCTGACTTAGGGAAGTCTGGAAAAATCTCCCTTTGGGGGG 96
DB 258 CCCAAGAACACAGCAACCACTGGGAAAGACAGGACACTCTCCAGCCTGGCACCC 199
QY 97 CAGGGGCGGGGATGAGCCAGGCGGAGAAAGAACTCTGAAGACTCGTAGATTGCTCTA 156
DB 198 TAAATCTGGAGGAGGAAGCTCTTGCTGGGGGGAACATTACAGAGCCTGGAGCATAGCTTA 139
QY 157 GACCGCTCAGACACTCTCTGGCGCAGCGTGTGAGAGGATTTGTGCA 201
DB 138 CACCTCATAGGGAAGACAGGCGCTGATCTGTGTGAGGTTGGCCA 94

RESULT 3
AAH15417/C
ID AAH15417 standard; CDNA; 1612 BP.
XX
AC AAH15417;
XX
XX 26-JUN-2001 (first entry)
DT
XX

DE Human cDNA sequence SEQ ID NO:13627.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 13627; 2537P + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide which comprises a 3'-end sequence, where the
XX polynucleotide comprises at least 15 nucleotides and the combination of
XX oligonucleotide which comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 1612 BP; 352 A; 476 C; 391 G; 393 T; 0 other;
XX
XX Query Match 18.4%; Score 38.6; DB 22; Length 1612;
XX Best Local Similarity 52.1%; Pred. No. 0.029; 79; Indels 0; Gaps 0;
XX Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
XX
XX 37 CCAGAAAGACAGCTGACTAGTGAAGTCTGGGAAAATCTCCTGCTTTGGGGGG 96
XX 258 CCCAAGAGACAGACGACCTCGGAAAGACAGGGACAGTCTCCAGCTGGGACCT 199
XX
XX 97 CAGGGGGGGGATGAGCCAGGCGGAGAGAACTCTGAAGACCTGAGATTGCTCTA 156
XX 198 TAAATCTGAGAGAGAGAGCTCTGCTGGGGAACTTACAGAGCTGAGAGATGATTA 139
XX
XX 157 GACCGCTCAGACACTCTCGGCGAGCGTGGAGAGATTGTGCA 201
XX 138 CACCTCAGGAGAGACAGCGCTGATCTGCTGGTGGGTTGCCCA 94
XX
XX Db

RESULT 4
ID AAL02939
ID AAL02939 standard; DNA; 13327 BP.
XX
XX AAL02939;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 5627.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX W020015320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184654.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226271.
XX 22-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227189.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.

PR	12-SEP-2000	2000US-0231968	PR	01-DEC-2000	2000US-0250391
PR	14-SEP-2000	2000US-0232397	PR	05-DEC-2000	2000US-0251030
PR	14-SEP-2000	2000US-0232398	PR	05-DEC-2000	2000US-0251030
PR	14-SEP-2000	2000US-0232399	PR	05-DEC-2000	2000US-0251030
PR	14-SEP-2000	2000US-0232400	PR	06-DEC-2000	2000US-0251479
PR	14-SEP-2000	2000US-0232401	PR	08-DEC-2000	2000US-0251479
PR	14-SEP-2000	2000US-0233063	PR	08-DEC-2000	2000US-0251856
PR	14-SEP-2000	2000US-0233064	PR	08-DEC-2000	2000US-0251856
PR	14-SEP-2000	2000US-0233065	PR	08-DEC-2000	2000US-0251856
PR	21-SEP-2000	2000US-0234223	PR	08-DEC-2000	2000US-0251856
PR	21-SEP-2000	2000US-0234274	PR	08-DEC-2000	2000US-0251856
PR	25-SEP-2000	2000US-0234997	PR	08-DEC-2000	2000US-0251856
PR	25-SEP-2000	2000US-0234998	PR	11-DEC-2000	2000US-0251990
PR	25-SEP-2000	2000US-0235484	PR	11-DEC-2000	2000US-0251990
PR	27-SEP-2000	2000US-0235834	PR	05-JAN-2001	2001US-0259678
PR	29-SEP-2000	2000US-0236327	XX		
PR	29-SEP-2000	2000US-0236367	XX		
PR	29-SEP-2000	2000US-0236368	XX		
PR	29-SEP-2000	2000US-0236369	XX		
PR	02-OCT-2000	2000US-0236370	XX		
PR	02-OCT-2000	2000US-0236802	XX		
PR	02-OCT-2000	2000US-0237037	XX		
PR	02-OCT-2000	2000US-0237038	XX		
PR	02-OCT-2000	2000US-0237039	XX		
PR	13-OCT-2000	2000US-0237040	XX		
PR	13-OCT-2000	2000US-0239935	XX		
PR	20-OCT-2000	2000US-0239937	XX		
PR	20-OCT-2000	2000US-0240960	XX		
PR	20-OCT-2000	2000US-0241221	XX		
PR	20-OCT-2000	2000US-0241785	XX		
PR	20-OCT-2000	2000US-0241786	XX		
PR	20-OCT-2000	2000US-0241787	XX		
PR	20-OCT-2000	2000US-0241808	XX		
PR	20-OCT-2000	2000US-0241809	XX		
PR	01-NOV-2000	2000US-0241826	XX		
PR	08-NOV-2000	2000US-0246474	XX		
PR	08-NOV-2000	2000US-0246475	XX		
PR	08-NOV-2000	2000US-0246476	XX		
PR	08-NOV-2000	2000US-0246477	XX		
PR	08-NOV-2000	2000US-0246478	XX		
PR	08-NOV-2000	2000US-0246523	XX		
PR	08-NOV-2000	2000US-0246524	XX		
PR	08-NOV-2000	2000US-0246525	XX		
PR	08-NOV-2000	2000US-0246526	XX		
PR	08-NOV-2000	2000US-0246527	XX		
PR	08-NOV-2000	2000US-0246528	XX		
PR	08-NOV-2000	2000US-0246532	XX		
PR	08-NOV-2000	2000US-0246609	XX		
PR	08-NOV-2000	2000US-0246610	XX		
PR	08-NOV-2000	2000US-0246611	XX		
PR	17-NOV-2000	2000US-0249207	XX		
PR	17-NOV-2000	2000US-0249208	XX		
PR	17-NOV-2000	2000US-0249209	XX		
PR	17-NOV-2000	2000US-0249210	XX		
PR	17-NOV-2000	2000US-0249211	XX		
PR	17-NOV-2000	2000US-0249212	XX		
PR	17-NOV-2000	2000US-0249213	XX		
PR	17-NOV-2000	2000US-0249214	XX		
PR	17-NOV-2000	2000US-0249215	XX		
PR	17-NOV-2000	2000US-0249216	XX		
PR	17-NOV-2000	2000US-0249217	XX		
PR	17-NOV-2000	2000US-0249218	XX		
PR	17-NOV-2000	2000US-0249244	XX		
PR	17-NOV-2000	2000US-0249245	XX		

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230439.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen
is used in preventing, treating or ameliorating a medical condition -
PS Disclosure; SEQ ID NO 9873; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.

CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 1838 BP; 404 A; 565 C; 468 G; 401 T; 0 other;

Query Match

Best Local Similarity 15.88; Score 33.2; DB 22; Length 1838;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 59 GCGAGCTCTGGGAAATCTCCCTCTTGGGGGCGAGCGCGGCGGATGACCAG 118
Db 816 GCGAGCGACGTGGGCAATGCGCCCTCTTGGGATGATGGTGTAAAGCCAG 757
Qy 119 GCGGAGAG 128
Db 756 GCGGAGAG 747

RESULT 10
AAL36532

ID AAL36532 standard; DNA: 7601 BP.

XX AAL36532;

DT 08-JAN-2002 (first entry)

DE Human musculoskeletal system related polynucleotide SEQ ID NO 2897.

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilice;
KW vunerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ds.
KW musculoskeletal system; ds.

XX Homo sapiens.

PN MO200155367-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001MO-US01338.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 02-MAR-2000; 2000US-0184664.
PR 16-MAR-2000; 2000US-0186350.
PR 17-MAR-2000; 2000US-0189874.
PR 18-MAR-2000; 2000US-0190076.
PR 19-MAR-2000; 2000US-0198123.
PR 07-JUN-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0214886.
PR 07-JUL-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246417.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254079.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

XX PA
 XX PI Rosen CA, Barash SC, Ruben SM;
 XX DR WPI; 2001-451937/48.

XX PT Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -

XX PS Example 2; SEQ ID NO 2897; 781bp + Sequence listing; English.

XX CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (AAB03087-AAB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7601 BP; 2275 A; 1450 C; 1551 G; 2325 T; 0 other;

Query Match 15.8%; Score 33.2; DB 22; Length 7601;
 Best Local Similarity 54.0%; Pred. No. 2.3; 58; Indels 0; Gaps 0;
 Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
 OY 9 TTCCTTCCTCCAGCAGCTGTGACAGAGCTCCAGAAAGCAGAGCTGACTTGGAGAGCTCT 68
 DB 6420 TTCATTTCTCTCACTGATGACATGAGAGATATATACAGAGATGCTAGGCTTGTCT 6479
 OY 69 GGGAAAAATCTCCCTGCTTTTGGGGGGGCGGCGGCGGATGAGCCAGGCGAGAGC 128
 DB 6480 GAGGAATTTATGGGTGTATTTGAGTACTGAGAGAGAGTGAGCGAGAGAGAGAAA 6539
 OY 129 AACTCT 134
 DB 6540 AACACT 6545

RESULT 11

AAL36533 standard; DNA; 7601 BP.

XX AAL36533;

XX 08-JAN-2002 (first entry)

DE Human musculoskeletal system related polynucleotide SEQ ID NO 2898.
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX antileptogenic; hepatotropic; antidiabetic; antileptogenic; antilept;
 XX vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; human; secreted protein;
 XX musculoskeletal system; ds.

XX Homo sapiens.

XX WO20015367-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220964.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225477.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235634.
 PR 27-SEP-2000; 2000US-0235636.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241781.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250300.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-451937/48.

Isolated polypeptide for treating, preventing and/or prognosing disorders related to the musculoskeletal system including musculoskeletal cancers and also for testing and detection e.g. diagnosis -

Example 2; SEQ ID NO 2898; 781bp + Sequence Listing; English.

XX The invention relates to novel genes (AAL34669-AAL37666) and proteins (ABB03087-ABB04109) associated with the musculoskeletal system useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, liver, lung, or urogenital; (b) gastrointestinal tract, liver, lung, or urogenital; (c) Crohn's disease, disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences.

XX

Sequence 7601 BP: 2275 A; 1450 C; 1551 G; 2325 T; 0 other;

Query Match 15.8%; Score 33.2; DB 22; Length 7601;
 Best Local Similarity 54.0%; Pred. No. 2.3;
 Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

PR	29-SEP-2000	2000US-0236367	
PR	29-SEP-2000	2000US-0236368	
PR	29-SEP-2000	2000US-0236369	
PR	29-SEP-2000	2000US-0236370	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237038	
PR	02-OCT-2000	2000US-0237039	
PR	02-OCT-2000	2000US-0237040	
PR	13-OCT-2000	2000US-0239935	
PR	13-OCT-2000	2000US-0239937	
PR	20-OCT-2000	2000US-0240360	
PR	20-OCT-2000	2000US-0241121	
PR	20-OCT-2000	2000US-0241185	
PR	20-OCT-2000	2000US-0241787	
PR	20-OCT-2000	2000US-0241808	
PR	20-OCT-2000	2000US-0241809	
PR	20-OCT-2000	2000US-0241826	
PR	01-NOV-2000	2000US-0244617	
PR	08-NOV-2000	2000US-0246474	
PR	08-NOV-2000	2000US-0246525	
PR	08-NOV-2000	2000US-0246537	
PR	08-NOV-2000	2000US-0246537	
PR	08-NOV-2000	2000US-0246558	
PR	08-NOV-2000	2000US-0246532	
PR	08-NOV-2000	2000US-0246609	
PR	08-NOV-2000	2000US-0246610	
PR	08-NOV-2000	2000US-0246611	
PR	17-NOV-2000	2000US-0249207	
PR	17-NOV-2000	2000US-0249208	
PR	17-NOV-2000	2000US-0249210	
PR	17-NOV-2000	2000US-0249211	
PR	17-NOV-2000	2000US-0249212	
PR	17-NOV-2000	2000US-0249213	
PR	17-NOV-2000	2000US-0249214	
PR	17-NOV-2000	2000US-0249215	
PR	17-NOV-2000	2000US-0249216	
PR	17-NOV-2000	2000US-0249217	
PR	17-NOV-2000	2000US-0249218	
PR	17-NOV-2000	2000US-0249244	
PR	17-NOV-2000	2000US-0249245	
PR	17-NOV-2000	2000US-0249264	
PR	17-NOV-2000	2000US-0249265	
PR	17-NOV-2000	2000US-0249297	
PR	17-NOV-2000	2000US-0249299	
PR	01-DEC-2000	2000US-0250160	
PR	01-DEC-2000	2000US-0250191	
PR	05-DEC-2000	2000US-0251988	
PR	05-DEC-2000	2000US-0251988	
PR	05-DEC-2000	2000US-0251988	
PR	06-DEC-2000	2000US-0251986	
PR	08-DEC-2000	2000US-0251865	
PR	08-DEC-2000	2000US-0251869	
PR	08-DEC-2000	2000US-0251890	
PR	11-DEC-2000	2000US-0254097	
PR	05-JAN-2001	2001US-0259678	
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Rosen CA, Barash SC, Ruben SM		
PI			
XX			

DR	WPI; 2001-465570/50.
XX	
PT	Isolated nucleic acid molecule encoding a reproductive system antigen
PT	is used in preventing, treating or ameliorating a medical condition -
PS	Disclosure: SEQ ID NO 7415; 1297pp + Sequence Listing; English.
XX	
CC	The present invention provides the protein and coding sequences of a
CC	number of human reproductive system related antigens. These can be used
CC	in the prevention and treatment of reproductive system disorders,
CC	including cancer. The present sequence is a genomic sequence encoding a
CC	protein of the invention.
SQ	Sequence 11360 BP; 2899 A; 2709 C; 2708 G; 3044 T; 0 other;
OY	Query Match 15.8%; Score 33.2; DB 22; Length 11360;
DB	Best Local Similarity 67.1%; Pred. No. 2.6;
MATCHES	Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
OY	59 GGGAGGCTGTGGAAAAATCTCCCTTGCTTTTGGGGGCGACGGCGGGGATGCAGCAGG 118
DB	4820 GGGAGGACTGTGGGAATGCCCTTCCTTGGGGGTGGAATAGGTGGGTATAAAGGCAAG 4761
OY	119 GCCGAGAAGG 128
DB	4760 GCCGGGAAGG 4751
RESULT 14	
ID	AAA94801/c
XX	AAA94801 standard; DNA; 553 BP.
XX	
AC	AAA94801;
DT	23-FEB-2001 (first entry)
DE	Drosophila melanogaster odorant receptor DOR62 coding sequence.
KW	odorant receptor; fruit fly; DOR62; odour recognition; pest control; ds.
OS	Drosophila melanogaster.
FH	Key Location/Qualifiers
FT	CDS 1..462
FT	/+tag= a
FT	/product= "DOR62"
FT	/partial
PN	WO2000050566-A2.
PD	31-AUG-2000.
PX	25-FEB-2000; 2000WO-US04995.
PX	25-FEB-1999; 99US-0257706.
PA	(UYCO) UNIV COLUMBIA NEW YORK.
PL	Vosshall LB, Amrein HO, Axel R;
DR	WPI; 2000-572081/53.
DR	P-PSDB; AAB26401.
PT	Novel nucleic acid encoding an insect odorant receptor, for identifying
PT	modulator compounds that are useful in controlling pest population -
PS	Claim 4; Fig 8; 176pp; English.
CC	The present sequence is the coding sequence for the Drosophila
CC	melanogaster odorant receptor DOR62. It was isolated using a differential
CC	cloning strategy. The odorant genes and proteins are useful as they aid
CC	in the study of the olfactory organ in mammals, as well as aiding the
CC	understanding of the link between odour recognition and behaviour in

CC Insects. They also enable the identification of compounds capable of
CC activating and inhibiting the receptors, allow the control of pest
CC populations via the use of alarm odour ligands and via the use of ligands
CC which interfere with the interaction between odorant ligands and
CC receptors associated with fertility.

sequence 553 BP; 141 A; 145 C; 132 G; 135 T; 0 other;

Query Match	15.4%	Score 32.4	DB 21	Length 553
Best Local Similarity	48.4%	Pred. No. 1.8		
Matches 90; Conservative	0;	Mismatches 96;	Indels 0;	Gaps 0;

QY	2	CGCCTTGTCTTTTTCCTCCACCCGTGCAAGAAATCCCAAGAAACACGCTAGTAAAGG	61
Db	264	CGCCTTCGCTCTGTCTCGCATCTGTCCCAAAATACAGATTACAAACACCTTCCAAGT	205
QY	62	AAGCTCTGGAAAAATCTCCCTGTTTGGGGGGGACAGCGGGGGGATGAGCCAGGGCC	121
Db	204	GAGCGCGGAGAAAAATACAAATCGAGATGATCATGCGCGGTGTGGCTGTGGTCATCCGCTAC	145
QY	122	GAGAGGAACCTGTGAAGACTCCGTAGATTGCTCTAGACCGCTTCAGACACTCTGGGGCA	181
Db	144	GTACACGAAATGCATGCGCAGCGGTACACTGGAGCGCGCGGAGCAGACGAACTGGGCAT	85
QY	182	GGCTGG	187
Db	84	GCAAGG	79

RESULT 15

AAK92113
ID AAK92773 standard; cDNA; 555 BP.
XX

AC AAK92773;

06-NOV-2001 (first entry)

Human CDNA 3'-end sequence, SEQ ID NO: 1233

human; full length cDNA; cDNA synthesis; oligo-capping; ss-

Homo sapiens.

PN EP1.130094-A2.

PD 05-SEP-2001

07-JUL-2000; 2000EP-0114089

PR 08-JUL-1999; 99JP-0194486
11-JAN-2000; 2000JE-0138774

02-MAY-2000; 20000JP-0183765

... (RECEIVED) / HELIX RES INST.

Wakamatsu A, Sugiyama

DR WPI; 2001-524255/58.

830 Primers useful for synthesizing full length cDNA clones and th

[illegible]

The invention relates to a method for determining the position of a mobile object relative to a reference point.

isolated and nucleotide sequences of cDNA molecules encoding a human protein have been

CDNA are useful for clarifying the function of the protein.

length enriched cDNA libraries that were synthesized by the

Sequence 555 BP; 124 A; 148 C; 167 G; 110 T; 6 other;
 CC Without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SO

50 Sequence 555 BP; 124 A; 148 C; 167 G; 110 T; 6 other,

Query Match	15.4%	Score 32.4	DB 22	Length 555
Best Local Similarity	49.4%	Pred. No. 1.8		
Matches 84; Conservative	0;	Mismatches 86;	Indels 0;	Gaps 0

QY 19 TCAGCCTGTCAAGGAAGTCCAGAAAGCAGCTGACTTTAGGGAAGGTCTGGGAAAAATC 78

Db 85 TTAAACCTGTAATTTAAAGCACACACAAACAAAAGGTGCATATGGGAATCTCGGGGCCAACGC 144

QY 79 TCCCTGCTTTGGGGGGGGCAGGGCGGGGGATGAGCCAGAGGCCAGAAAGAACTCTTAAG 133

Db 145 CCCACGATGTAGGGGAGGCCCTGATTCGATCGGGGGCGGGGGCTTTCTCCATGGGGCTCG 204

QY 139 ACTCCGTAGATTGCTTAGACCGGCTCAGACATCTCGGGGAGACGTGGA 188

Db 205 GCCCCTTAGTGAAGGTGTACACTCTCCACAGATCAGAGAGCTTCTTGGA 254

```
Search completed: November 6, 2002, 14:17:11
Job time : 63.7914 secs
```



```

/sex="Female"
/lab.host="E. coli strain XL10-Gold, T1-resistant, F-"/
/Note="Vector: PMD29v: Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD2 (91473211419b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      173 a      189 c      147 g      180 t
ORIGIN

```

```

Query Match      40.2% Score 667.8; DB 12; Length 689;
Best Local Similarity 99.7%; Pred. No. 4.2e-134;
Matches 669; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 924 AACCCCACTATTTCTCAGCCCTCAGATGAAGAAGTATGTTGTTGAGAGATGATG 993
    |||||||
DB 689 AACCCCACTATTTCTCAGCCCTCAGATGAAGAAGTATGTTGTTGAGAGATGATG 930

QY 984 GGTTCAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1043
    |||||||
DB 629 GGTTCAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 570

QY 1044 TTGTTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1103
    |||||||
DB 569 TTGTTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 510

QY 1104 GTGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1163
    |||||||
DB 509 GTGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 450

QY 1164 TGCTTCCATCCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1223
    |||||||
DB 449 TGCTTCCATCCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 390

QY 1224 GCTTTGCTGAGACCTGGGCTGAGAGTGTGTTGTTGAGAGAGGAAACCTCATCTTG 1283
    |||||||
DB 389 GCTTTGCTGAGACCTGGGCTGAGAGTGTGTTGTTGAGAGAGGAAACCTCATCTTG 330

QY 1284 TTACTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1343
    |||||||
DB 329 TTACTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 270

QY 1344 AAAAGTTGATGTTTGAAGAAATCGTGCTTAAAGAAACCTCACTGAGATGGGGG 1403
    |||||||
DB 269 AAAAGTTGATGTTTGAAGAAATCGTGCTTAAAGAAACCTCACTGAGATGGGGG 210

QY 1404 CCATCTCTCCAGACCTCGAATCTCAATCTGCTGCTGCTGAAGATGAATGAATCTCTGG 1463
    |||||||
DB 209 CCATCTCTCCAGACCTCGAATCTCAATCTGCTGCTGCTGAAGATGAATGAATCTCTGG 150

QY 1464 AATGTAACCATGCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1523
    |||||||
DB 149 AATGTAACCATGCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90

QY 1524 ACGCTTTGGGGTGAAGTCTGAGAAATCTCCAGATTTTGAAGAGGGGAGGGGAGGGA 1583
    |||||||
DB 89 ACGCTTTGGGGTGAAGTCTGAGAAATCTCCAGATTTTGAAGAGGGGAGGGGAGGGA 30

QY 1584 TGAGAAACTTG 1594

```

Db 29 TGAGAAACTG 19

RESULT 2
LOCUS
DEFINITION

CNS03SBM 946 bp DNA linear GSS 17-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
051620 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL258341.1 GI:7979353
VERSION AL258341.1
KEYWORDS GSS; genome survey sequence.
SOURCE
ORGANISM Tetraodon nigroviridis

REFERENCE
AUTHORS Reest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

TITLE
JOURNAL Unpublished
AUTHORS Reest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

REFERENCE
AUTHORS Reest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

TITLE
JOURNAL Unpublished
AUTHORS Reest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

TITLE
JOURNAL Unpublished
AUTHORS Reest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
1. .946
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_id="051620"
/clone_lib="G"
/note="Genoscope sequence ID : COB6051BD10LP1-end : T7"

BASE COUNT 191 a 291 c 305 g 133 t 26 others
ORIGIN

Query Match 4.2% Score 70.4; DB 12; Length 946;
Best Local Similarity 67.9%; Pred. No. 5.5e-05;
Matches 95; Conservative 2; Mismatches 43; Indels 0; Gaps 0;

QY 155 GGTGCTTTCGCCACATGCGAGCCGAGCTCCATGATGAGACAGATAGCTTTA 214
 |||||
DB 707 GGTGCTTTCGCCACATGCGAGCCGAGCTCCATGATGAGACAGATAGCTTTA 766

QY 215 CAGCTCATTTGAAAAC 274
 |||||
DB 767 GAGCTCCACCCCAAC 826

QY 275 ACACATGTTTCTCCCTCC 294
 |||||
DB 827 ACATATGAGTCACTCCTCC 846

RESULT 3
CNS02GTM/C 681 bp DNA linear GSS 13-MAY-2000
LOCUS
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
138622 of library G from Tetraodon nigroviridis, genomic survey

ACCESSION		sequence.
VERSION	AJ196781	
KEYWORDS	GSS; genome survey sequence.	
SOURCE	Tetraodon nigroviridis.	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	
REFERENCE	1 (bases 1 to 681)	
AUTHORS	Ruest-Crollius,H., Jalllon,O., Dasilva,C., Filames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot.A. and Weissenbach,J.	
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 681)	
AUTHORS	Ruest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Filames,C., Wincker,P., Brotlier,P., Queller.F., Saurin,W. and Weissenbach,J.	
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 681)	
AUTHORS	Direct Submission	
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases	
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.	
FEATURES	Location/Qualifiers	
Source	1..681 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /collapse="138G22" /clone.lib="G" /note="Genoscope sequence ID : COAG138BD11SP1-end ; PUC-Ori"	
BASE COUNT	199 a 121 c 169 g 166 t 26 others	
ORIGIN		
Query Match	4.0%; Score 66.8; DB 12; Length 681;	
Matches Local Similarity	81.9%: Pred. NO.0.00033;	
Matches 77: Conservative	0; Mismatch 17; Indels 0; Gaps 0;	
Oy	229 ACACACACACACACACACACACACACACACACACACATGTTCCTT 288 	
Db	569 ACACACACACAAACAACACACACACACACACACACACACTTGCGCT 530 	
Oy	289 CCCTCCCTCACTTCGCCATTCTGTGGTGC 322 	
Db	529 CCTCCTCCCCTCCGCTCATTTGCTTTGGTGC 496 	
RESULT 4		
LOCUS	AZ463187	
DEFINITION	IM0271B21R Mouse 10kb plasmid UDCGM library Mus musculus genomic clone UDCGM0271B21 R, DNA sequence.	
ACCESION	AZ463187	
VERSION	AZ463187.1 GI:10621312	
KEYWORDS	GSS.	
SOURCE	house mouse. Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 511) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill.C., Islami.H., Longacre,S., Mahmoud.M., Meenen.E., Pedersen.T., Reilly ,M., Rose,M., Rose,R., Stokes.R., Tingey,A., von Niederhausern,A. and Wright.D.,Weiss,R.	

TITLE	JOURNAL	COMMENT	FEATURES
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	Unpublished (2000)	Contact: Robert B. Weiss University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: rdunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0271 row: B column: 21 Seq primer: CACACGAGAAACAGCTATGACC Class: plasmid ends High quality sequence stop: 511.	Location/Qualifiers 1. 511 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="U061M0271B21" /clone_11b="Mouse 10kb plasmid U061M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-" /note="Vector: pMD2429; Purified genomic DNA from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g11473211419b1AFL29072.1) a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT	139 a	165 c	78 g 129 t
ORIGIN			
Query Match	4.0%; Score 66.2; DB 12; Length 511;		
Best Local Similarity	68.1%; Pred. No. 0.00045;		
Matches	92; Conservative	0; Mismatches	43; Indels 0; Gaps 0;
QY	181	CGGAGCTCCAGTCGATATGACAGAAATAGCTTACAGCTACATTCMAAACACACACACAC	240
DB	246	CAGGAGAGATAGCTTATGTTCTTCTGCGCTGCTTCACACAACCTACACACACACAC	305
QY	241	ACACACACACACACACACACACACACACACACATGTTTCTTCCTCCCTCCAC	300
DB	306	ACACACACACACACACACACACACACACACACATGAGCCATCATCCTTC	365
QY	301	TTCCCTCCATCTCTC	315
DB	366	TGCTTGGCGTTTCT	380
RESULT 5			
CNS038VZ/c	958 bp DNA linear GSS 15-MAY-2000		
LOCUS	Tetradon nigroviridis genome survey sequence T7 end of clone		
DEFINITION	006A06 of library G from Tetradon nigroviridis, genomic survey sequence.		
ACCESSION	AL233144		
VERSION	AL233144.1 GI:7892279		
KEYWORDS	GSS; genome survey sequence.		

SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 958)
AUTHORS Roest-Crolius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 958)
AUTHORS Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Saurin,W., Fizames,C., Wincker,P., Brottier,P., Queller,F., Bernot,A. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 958)
AUTHORS Genoscope.
TITLE Direct Submission
REFERENCE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.
FEATURES
source
1. 958
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="006A06"
/clone_1lb="G"
/note="Genoscope sequence ID : C05G006BA03LPI-end : T7"
BASE COUNT 247 a 202 c 229 g 276 t 4 others
ORIGIN
Query Match 3.9%; Score 65.4; DB 12; Length 958;
Best Local Similarity 63.9%; Pred. No. 0.00066;
Matches 99; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 131 AAGATGGAATATTTGTACAAAGAGGCTTTCTCCCGACCATGCGACCGGAGCTCC 190
Db 701 ATCAAGAGAAATCTACGATGACAGTTGATACCTGTCCAGAGGGGAGCAGCAGCGGAC 642
QY 191 ACATGATATGACACAAATAGCTTTACAGCTACATTCAAAACACACACACACACAC 250
Db 641 ACAGATTAACCCACAGATTAACGACAGCTTGTATATACACACACACACACACAC 582
QY 251 ACACACACACACACACACACACACACATGTTT 285
Db 581 ACACACACACACACACACACACACACACGCTT 547

RESULT 6
A2394294 684 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0157J1BR Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0157J18 R, DNA sequence.
ACCESSION A2394294
VERSION A2394294.1 GI:10509366
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 684)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Plasmid Inserts
COMMENT Unpublished (2000)
CONTACT: Robert B. Weiss
UNIVERSITY OF UTAH Genome Center
UNIVERSITY OF UTAH
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
TEL: 801 585 5606
FAX: 801 585 7177
EMAIL: ddunn@genetics.utah.edu
INSERT LENGTH: 10000 Std Error: 0.00
PLATE: 0157 row: 7 column: 18
SEQ PRIMER: CACACAGAAACACGCTATGACC
CLASS: plasmid ends
HIGH QUALITY SEQUENCE STOP: 684.
FEATURES
source
1. 684
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0157J18"
/clone_1lb="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1473214[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 212 a 159 c 133 g 180 t
ORIGIN
Query Match 3.9%; Score 65.2; DB 12; Length 684;
Best Local Similarity 69.8%; Pred. No. 0.00073;
Matches 88; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 155 GGTGCTTTCCTCCCGACGAGCGGAGGCTCATGATGACAGAAATGACTTA 214
Db 145 GGTTCATCTCTGTACACATAAAGTGGGACATACAGACTTGTCTCAAAAGCTTAA 204
QY 215 CAGCTACATTCAAACACACACACACACACACACACACACACACACACACAC 274
Db 205 AAGCTCATTAAC 264
QY 275 ACACAT 280
Db 265 ACACAT 270

RESULT 7
CNS020RH 729 bp DNA linear GSS 12-MAY-2000
LOCUS CNS020RH
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 224G15 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL175958
VERSION AL175958.1 GI:7814015
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.

QY 287 TTCCCTCCCTCCACT 301
||||| | | |||
Db 423 TTCCGACAAATGTTACT 409

Search completed: November 6, 2002, 16:06:59
Job time : 2070.91 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:20:41 ; Search time 285.578 Seconds
(without alignments) 15388.369 Million cell updates/sec

Title: US-09-833-740-7_COPY_1_210

Perfect score: 210

Sequence: 1 ccgcctgtttcttctctc.....ggatttgcaaacattcc 210

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgtg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	210	100.0	527	6	AX354813	AX354813 Sequence
2	210	100.0	198271	2	AC026591	AC026591 Homo sapi
3	210	100.0	210719	2	AC087646	AC087646 Homo sapi
4	210	100.0	213359	2	AC069006	AC069006 Homo sapi
5	62.2	29.6	2170	6	AX354807	AX354807 Sequence
6	62.2	29.6	3130	10	AF338224	AF338224 Mus muscu
7	62.2	29.6	207160	2	AC016464	AC016464 Mus muscu
8	62.2	29.6	211173	2	AL646097	AL646097 Mus muscu
9	54	25.7	493	6	AX354812	AX354812 Sequence
10	38.6	18.4	1612	2	AK021842	AK021842 Homo sapi
11	38.6	18.4	202682	2	AL513327	AL513327 Homo sapi
12	36.8	17.5	147668	2	AC106475	AC106475 Rattus no
13	35.8	17.0	145617	2	AC096681	AC096681 Rattus fam
14	35.8	17.0	148198	2	AC024232	AC024232 Homo sapi
15	35.8	17.0	176831	2	AC084337	AC084337 Homo sapi
16	35.6	17.0	153394	2	AC092731	AC092731 Felis cat
17	35.6	17.0	153020	2	AF429315	AF429315 Homo sapi
18	35.2	16.8	125020	9	AC008755	AC008755 Homo sapi
19	35.2	16.8	203912	9	AC008755	AC008755 Homo sapi
20	34.6	16.5	105288	2	AC084737	AC084737 Homo sapi
21	34.6	16.5	145598	9	AC008132	AC008132 Homo sapi
22	34.6	16.5	150724	9	AP000550	AP000550 Homo sapi
23	34.6	16.5	150754	9	AC023491	AC023491 Homo sapi
24	34.6	16.5	159550	2	AC013360	AC013360 Homo sapi
25	34.6	16.5	163795	2	AP000356	AP000356 Homo sapi
26	34.6	16.5	165050	9	AC007325	AC007325 Homo sapi
27	34.6	16.5	170667	2	AC093573	AC093573 Pan trogl
28	34.6	16.5	175358	9	AC007981	AC007981 Homo sapi
29	34.6	16.5	180884	9	AC008018	AC008018 Homo sapi
30	34.2	16.3	64836	2	AC102520	AC102520 Mus muscu
31	34.2	16.3	191656	2	AC068733	AC068733 Mus muscu
32	34.2	16.3	214194	2	AC073757	AC073757 Mus muscu
33	34.2	16.3	233448	2	AC073692	AC073692 Mus muscu
34	34.2	16.3	237613	2	AC025769	AC025769 Homo sapi
35	34	16.2	197426	3	CNS07EG9	AL590442 Chromosom
36	33.8	16.1	148054	9	AC026467	AC026467 Homo sapi
37	33.6	16.0	141990	9	AC004691	AC004691 Homo sapi
38	33.4	15.9	1368	1	PSEFUDH	D32042 Pseudomonas
39	33.4	15.9	62875	2	AC068436	AC068436 Homo sapi
40	33.4	15.9	110000	2	HS1171M.2	Continuation (3 of
41	33.4	15.9	118229	2	AC013315	AC013315 Homo sapi
42	33.4	15.9	148539	2	AC094804	AC094804 Rattus no
43	33.4	15.9	177990	2	AC016385	AC016385 Homo sapi
44	33.4	15.9	205802	2	AC015802	AC015802 Homo sapi
45	33.4	15.9	340000	9	HS21C102	AL163302 Homo sapi

ALIGNMENTS

RESULT 1	AX354813	527 bp	DNA	Linear	PAT 06-FEB-2002
LOCUS	AX354813	Sequence	7 from Patent WO0179290.		
DEFINITION	AX354813				
ACCESSION	AX354813.1	GI:18619545			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
AUTHORS
Drucker,D.J. and Lovshin,J.A.
TITLE
Glp-2 receptor gene promoter and uses thereof
JOURNAL
Patent: WO 0179290-A 7 25-Oct-2001;
1149336 ONTARIO INC. (CA)
FEATURES
source
Location/Qualifiers
1..527
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	111 a	140 c	163 g	113 t
------------	-------	-------	-------	-------

AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213359)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Dec 11, 2000 this sequence version replaced gi:9280811.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0773021
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 206822 bases at least Q40
Consensus quality: 208592 bases at least Q30
Insert size: 214000; agarose-fp
Insert size: 211859; sum-of-contigs
Quality coverage: 5.34 in Q20 bases; agarose-fp
Quality coverage: 5.44 in Q20 bases; sum-of-contigs

----- NOTE: This is a working draft sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----

1 1400: contig of 1400 bp in length
1401 1500: gap of unknown length
1501 4621: contig of 3121 bp in length
4622 4721: gap of unknown length
4722 7894: contig of 3173 bp in length
7895 7994: gap of unknown length
7995 10806: contig of 2812 bp in length
10807 10906: gap of unknown length
10907 16818: contig of 5912 bp in length
16819 16918: gap of unknown length
16919 24662: contig of 7744 bp in length
24663 24762: gap of unknown length
24763 32389: contig of 7627 bp in length
32390 32490: gap of unknown length
32490 40414: contig of 7925 bp in length
40415 40514: gap of unknown length
40515 53087: contig of 12573 bp in length
53088 53187: gap of unknown length
53188 64676: contig of 11489 bp in length
64677 64776: gap of unknown length
64777 78536: contig of 13760 bp in length
78537 78636: gap of unknown length
78637 94647: contig of 16011 bp in length
94648 94747: gap of unknown length
94748 117462: contig of 22715 bp in length
117463 143519: contig of 25957 bp in length
143520 143620: gap of unknown length
143620 171176: contig of 27357 bp in length
171177 213359: contig of 42083 bp in length.

Location/Qualifiers
1..213359
/organism="Homo sapiens"

/db_xref="taxon:9606"
/chromosome="RP11-773021"
/clone="RP11-773021"
1..1400
/note="assembly_name:Contig18"
1501..4621
/note="assembly_name:Contig19"
4722..7894
/note="assembly_name:Contig20"
7995..10806
/note="assembly_name:Contig21"
10907..16818
/note="assembly_name:Contig22"
16919..24662
/note="assembly_name:Contig23
clone_end:77
vector_side:right"
24763..32389
/note="assembly_name:Contig24"
32490..40414
/note="assembly_name:Contig25"
40515..53087
/note="assembly_name:Contig26"
53188..64676
/note="assembly_name:Contig27"
64777..78536
/note="assembly_name:Contig28"
78637..94647
/note="assembly_name:Contig29"
94748..117462
/note="assembly_name:Contig30"
117563..143519
/note="assembly_name:Contig31
clone_end:SP6
vector_side:left"
143620..171176
/note="assembly_name:Contig32"
171277..213359
/note="assembly_name:Contig33"

BASE COUNT 59337 a 47447 c 48595 g 56470 t 1506 others

ORIGIN

Query Match 100.0%; Score 210; DB 2; Length 213359;
Best Local Similarity 100.0%; Pred. No. 7.5e-54;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCTTGTCTTCTCTCTCAAGCTGCAAGAAAGTCCGAAAGACACAGCTGACTTAGG 60
|||||
DB 83670 CCGCTTGTCTTCTCTCTCAAGCTGCAAGAAAGTCCGAAAGACACAGCTGACTTAGG 83611

QY 61 GAAGGCTGTGGAAATATCTCCGCTTTGGGGGGGCGAGGGGGATGAGCGAGGCG 120
|||||
DB 83610 GAAGGCTGTGGAAATATCTCCGCTTTGGGGGGGCGAGGGGGATGAGCGAGGCG 83551

QY 121 CGAAGAAAGAACTGGAAGACTCCGTAGATTGCTTAGACCGCTCAGACACTCTCGGGCG 180
|||||
DB 83550 CGAAGAAAGAACTGGAAGACTCCGTAGATTGCTTAGACCGCTCAGACACTCTCGGGCG 83491

QY 181 AGCGTGAGAGAGATTGTGCAAACTTTCC 210
|||||
DB 83490 AGCGTGAGAGATTGTGCAAACTTTCC 83461

RESULT 5
AX354807
LOCUS AX354807 2170 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0119290.
ACCESSION AX354807
VERSION AX354807.1 GI:18619538
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.

REFERENCE 1 (sites)
AUTHORS Drucker,D.J. and Losvshin,J.A.
TITLE G1P-2 receptor gene promoter and uses thereof
JOURNAL Patent: WO 0119290-A 1 25-OCR-2001;
110016 Cambridge, MA 02142

BASE COUNT	/note="Recombinant DNA expression construct"				
574	a	500	c	551	g
				540	t
ORIGIN				5	others

Query Match	29.6%;	Score 62.2;	DB 6;	Length 2170;
Best Local Similarity	67.5%;	Pred. No. 1.9e-08;		
Matches 141;	Conservative 0;	Mismatches 48;	Indels 20;	Gaps 3

0Y 3 GCGTTGTTCTTCTCCAGCGTCTCAAGGAAGTCCAGAAACAGCAGCTGACTT-AGGG 6L
1 1111 1111 11 11 11111111 11 11 111 111
Db 1476 GTCCTTGCTTTCTCTCGGCTTCTGAGGAAGTCCAGGCAGCGTGTGAGCTCTGGGG 153
0Y 62 AAGGCTGGGAAATCTCCCTGCTTTTGGGGGGGACAGGGGCGGGGANTAGCCAGGGCC 121

```

Oy      122  GAGAGGAACCTCTGAGACCTCCGTAGATTGCTCTAGACCGGCTCAGACACTCTCGGCCA 181
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1587  -----GAAACTTGGAGATTCCGTAATCCGCTAGAGCAACTCAGACAGTC----GGCG 1636

```

Oy	182 GCGTGAGAGGATTTGTCAACATTTC	210
Db	1637 GCCTGAAGAGACTTGTGCACAACACTTCC	1665

LOCUS	3130 bp	DNA	linear	ROD 18-JUN-2001
DEFINITION	Mus musculus glucagon-like peptide-2 receptor gene, partial cds.			
ACCESSION	AF338224			
VERSION	AF338224.1	GI:14485646		

REFERENCE AUTHORS TITLE JOURNAL	
1 (pages 1 to 3130) Lovshin, J.A., Estrill, J., Yusta, B., Brown, T.J. and Drucker, D.J. Glucagon-like peptide (GLP)-2 action in the Murine Central Nervous System is Enhanced by Elimination of GLP-1 Receptor Signaling J. Biol. Chem. 276 (24), 21489-21499 (2001)	

REFERENCE	2 (bases 1 to 3130)
AUTHORS	Lovshin, J. A. and Drucker, D. J.

JOURNAL Submitted (17-JAN-2001) Department of Medicine, University of Toronto, 101 College Street CCRW-845, Toronto, ON M5G2C4, Canada

FEATURES Location/Qualifiers

SOURCE 1. .3130

ORIGIN

Query Match	29.6%	Score 62.2;	DB 10;	Length 3130;
Best Local Similarity	67.5%	Pred. No. 1.9e+08;		
Matches 141;	Conservative	0;	Mismatches 48;	Indels 20;
				Gaps 3.

Qy 3 GCGCTGTTCTTCTCTCAGCCGTGTCAAGGAAGTCCCGAAMACACAGCTGACTT-AGGG 61
| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 2436 GTCTTGCTTTTTCTCTGCGCTTGCTGAGGAAGTCCCGACAGCAGTGTGTTGGGG 249

```

QY      62  AAGCTCTGGGAAAAATCTCCCTGCTTTTGGGGGGGACGGGGGGGGATGAGCCAGGGCC 121
          |||||
Db      2496 TAGTGCTGGGAAAAATCTCCCAAGATTTTAGAGGGGCGAGCGGGGGGATGA----- 2546

```

```

QY      122  GAGAAGAACTCTGAAGACTCCGTAGATTGCTCTAGACCGCCTCAGACACTCTCGGCCGA 181
          ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2547  -----GAAACTTGGAGATTCCGTAGATCCGCTAGAGCAACTCAGACAGTTC----GGCG 2596

```

Qy	182	GC	GT	GA	GA	GG	AT	TT	GT	GC	AA	CA	AT	TT	CC	210
Db	2597	GC	CT	GA	GA	GA	GA	CA	CT	GT	GC	AA	CA	CA	CT	2625

RESULT 7				
AC016464	AC016464	207160 bp	DNA	linear
LOCUS				
DEFINITION	Mus musculus chromosome 11 clone RP23-409J21 map 11,			WORKING DRAFT
ACCESSION	AC016464			
VERSION	AC016464.3	GI:7137319		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS	
1 (bases 1 to 207160)	
Birren, B., Linton, L., Nusbaum, C. and Lander, E.	
Mus musculus chromosome 11, clone RP23-409J21	
Unpublished	
2 (bases 1 to 207160)	
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,	

Baldwin, J., Barina, N., Beckert, R., Boguski, S., Collins, S., Collamore, A., Brown, A., Castile, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Deatrelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galand, J., Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karkats, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., McDonald, P., Margulis, N., McEwan, P., McBurn, A., McKernan, K., McLaughlin, J., Melnick, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severly, P., Stange-Thomann, N., Stojanovic, N., Sudpanthanian, A., Talmes, J., Tistayes, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wymann, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 1, 2000 this sequence version replaced gi:16970326.
This sequence version is identical to the previous one. BenoitMekler: 211

ACCESSION	AL646097
VERSION	AL646097.5
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	house mouse.
ORGANISM	Mus musculus.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Mammalia:	Eutheria:	Rodentia:	Sciurognathi:	Muridae: Murinae: Mus
1 (sites)	McLay, K.	Direct Submission	Submitted (08-JAN-2002)	Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On Jan 11, 2002 this sequence version replaced gi:18072579.
Center:	Wellcome Trust Sanger Institute	Genome Center		
Center code:	SC			
Web site:	http://www.sanger.ac.uk			
Contact:	humquerry@sanger.ac.uk			
Project Information				
Center project name:	BM338M9			
----- Summary Statistics				
Assembly program:	XSAP4: Version 4.5			
Sequencing vector:	plasmid; L08752; 100% of reads			
Chemistry:	Dye-terminator Big Dye; 100% of reads			
Consensus quality:	206626 bases at least Q40			
Consensus quality:	207698 bases at least Q30			
Consensus quality:	208567 bases at least Q20			
Insert size:	209573; sum-of-contigs			
Insert size:	207120; 2.8% error; agarose-fp			
Quality coverage:	9.05x in Q20 bases; sum-of-contigs quality			
coverage:	9.26x in Q20 bases; agarose-fp			

* NOTE: This is a 'working draft' sequence.				
* This record will be updated with the finished sequence				
* as soon as it is available and the accession number will				
* be preserved.				
Location/Qualifiers				
1. 211173				
/organism="Mus musculus"				
/db_xref="taxon:10090"				
/chromosome="11"				
/clone="RP23-338M9"				
/clone_1lb="RP23-23"				
1. 8490				
/note="assembly-fragment:04365				
fragment_chain:1"				
8591. 10949				
/note="assembly-fragment:02882				
fragment_chain:1"				
11050. 22821				
/note="assembly-fragment:01376				
fragment_chain:1"				
22922. 39369				
/note="assembly-fragment:01658				
fragment_chain:1"				
39470. 47112				
/note="assembly-fragment:00263				
fragment_chain:1"				
47213. 51333				
/note="assembly-fragment:02994				
fragment_chain:2"				
51434. 56653				
/note="assembly-fragment:00672				
fragment_chain:2"				
56754. 71877				
/note="assembly-fragment:03096				
fragment_chain:2"				
71978. 112160				
/note="assembly-fragment:01861				
fragment_chain:2"				
11261. 129027				
/note="assembly-fragment:01143				
fragment_chain:2"				
129128. 142170				
/note="assembly-fragment:00621				
fragment_chain:2"				
142271. 146192				
/note="assembly-fragment:00072				

[illegible]

```

Matches 129: Conservative 0; Mismatches 60; Indels 19; Gaps 2;

OY 3 GCCTGTTCTTCTCCAGCCTGTCAGAACTCCAGAAACAGCAGCTGACTTAAAGGA 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 GCTTGTGCTTTTCTTCTGCGCTTCTGAGAAAGTCCAGAGCAGCTGAGCTCTGGGGG 62
OY 63 AGCTCTGGAAATATCTCCCTGCTTTTGGGGGCGAGGCGGGGATGAGCCAGGGCCG 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 TAGGCTGGGAAATATCTCCAGAGATTTAGAGAGGGGCGAGGGGATGA----- 112
OY 123 AGAAGCACTCTGAGACCTCCGTAGATGCTCTAGACCCGCTTACAGACTCTGGCGGAG 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 113 -----GAAACTTGAGATTTGGTAGATCGCTGTAGAGCAACTCAGACAGTCG----GCGG 163
OY 183 CGTGAGAGAGGATTTGTGCAAAACATTTCC 210
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 CCGAAGAGGACTTTGTGCAAAACACTTCC 191

RESULT 10
AK021842/c
LOCUS
DEFINITION Homo sapiens CDNA FLJ11780 fls, clone HEMBA1005931, weakly similar
to ZINC FINGER PROTEIN 83.
ACCESSION AK021842
VERSION AK021842.1 GI:10433110
KEYWORDS oligo capping, fls (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head CDNA to
Homo, clone.lib:HEMBA1 clone:HEMBA1005931.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Matsutsumi,M., Hosokawa,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Makatsuki,A.,
Nakamura,Y., Nagahara,K., Masuhara,Y., Niimura,Y., and Iwayanagi,T.
NEDO human CDNA sequencing project
TITLE Unpublished (2000)
JOURNAL 2 (phases 1 to 1612)
REFERENCE Isogai,T. and Otsuki,T.
AUTHORS Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human CDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: CDNA full insert
sequencing; Research Association for Biotechnology: CDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
    Location/Qualifiers
        1..1612
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HEMBA1005931"
            /dev_stage="embryo, 10 weeks"
            /tissue_type="whole embryo, mainly head"
            /note="cloning vector: pME185FL3"

BASE COUNT 352 a 476 c 391 g 393 t

Query Match 18.4%; Score 38.6; DB 9; Length 1612;
Best Local Similarity 52.1%; Pred. NO. 0.34;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

37. CCCAGAAAGCAGCTGAGGAGAGCTCGGAAATATCTCCGCTTTGGGGGGG 96

```

```

Db 258 CCCAAGACACAGAGGACCACTGGGAAAGAGGAGACGCTCCACAGCTGGGACCCCT 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 97 CAGGGGCGGGGATGAGCCAGGCGGAGAGAGATCTGAAAGATCCGTAGATTCTCTA 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 TAAATCTGGAGAGAGAGAGCTTTGCTGGGGGAGACATTACAGAGCTGAGCATAGTTA 139
OY 157 GACCGGCTCAGACATCTCCGCGGAGCGCTGAGAGAGATTGTGCA 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 CACCTCATAGGAGAGAGAGGCTGATCCTGGTGGAGGTTTCCCA 94

RESULT 11
AL513327/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-415J8, *** SEQUENCING IN
PROGRESS ***; 3 unordered pieces.
ACCESSION AL513327
VERSION AL513327.22 GI:18476646
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Matthews,N.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humberq@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced g1:17384473.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humberq@sanger.ac.uk
----- Project Information
Center project name: BA1508
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 202251 bases at least Q40
Consensus quality: 202335 bases at least Q30
Insert size: 202482; sum-of-contigs
Insert size: 196528; 5.3% error; agarose-fp
Quality coverage: 10.24x in Q20 bases; sum-of-contigs Quality
coverage: 10.51x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 53967: contig of 53967 bp in length
* 53968 54067: gap of 100 bp
* 54068 147217: contig of 93150 bp in length
* 147218 147317: gap of 100 bp
* 147318 202682: contig of 55365 bp in length.
    Location/Qualifiers
        1..202682
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="RP11-415J8"
            /clone.lib="RPC1-11.2"
            1..53967
                /note="assembly fragment:03338
                fragment_chain:1"

```



```

misc_feature      54068..147217
                  /note="assembly_fragment:06038
                  fragment_chain:1"
misc_feature      147318..202682
                  /note="assembly_fragment:05678
                  fragment_chain:1"
BASE COUNT      52289 a 47427 c 48564 g 54202 t 200 others
ORIGIN
Query Match      18.4%; Score 38.6; DB 2; Length 202682;
Best Local Similarity 52.1%; Pred. No. 0.33;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 37 CCCAGAACACACAGCTGACTAGTACAGAGCTCTGTGAAAAATCTCCCTGTTGGGGGG 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78763 CCCAGAACACACAGAGCCACCTGTGGAAACAGAGGAGACATCTCCACGCTGGGACCC 78704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 97 CAGGGCGGGGGATGAGCCAGAGCCGAGAAAGAACTCTGAAAGCTCCGTAGATTGCTTA 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78703 TAAATCTGAGAGGAGAAAGCTCTGCTGGGGGAAACATTAACAGAGCTGGAGCATAGGTTA 78644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 157 GACCGCTCAGACACTCTGTGGCGCCAGCTGAGAGCATTTGTGCA 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78643 CACCTCATAGGAGAAAGAGCCCTGATCTGTGTGAGGTTTGCCCA 78599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AC106475      147868 bp      DNA      linear      HMG 12-JAN-2002
LOCUS      Rattus norvegicus clone CH230-160H18, *** SEQUENCING IN PROGRESS
DEFINITION      *** 75 unordered pieces.
ACCESSION      AC106475
VERSION      AC106475.1 GI:18138997
KEYWORDS      HTG: HMG5-PHASE1.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE      1 (bases 1 to 147868)
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
              Alstrooms,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barberia,J.,
              Benton,J., Bimagne,K., Blankenburg,K., Bonnin,D., Bouck,J.,
              Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
              Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Cartron,T.F.,
              Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
              Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
              Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
              Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
              Demu,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
              Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
              Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
              Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
              Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
              Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
              Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
              Hollins,B., Homsif,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
              Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
              Jonathan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
              Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
              Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
              Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
              Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
              Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
              Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
              Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
              Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S.,
              Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
              Peery,J., Perez,L., Peters,L., Pichens,R., Primus,E., Pu,L.L.,
              Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
              Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitani,N.,
              Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
              Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

```

```

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,C., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 147868)
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GK2D
Center clone name: CH230-160H18
----- Summary Statistics
findPhrapList
Assembly program: Phrap; version 0.990329First call to
Consensus quality: 102886 bases at least Q40
Consensus quality: 113109 bases at least Q30
Consensus quality: 121647 bases at least Q20
Estimated insert size: 94980; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agatose-fp estimation
Quality coverage: 1.1x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
8024      8023: contig of 8023 bp in length
8124      8123: gap of unknown length
81676     11675: contig of 3552 bp in length
11776     11775: gap of unknown length
11776     15149: contig of 3374 bp in length
11776     15249: gap of unknown length
15150     19174: contig of 3925 bp in length
15250     19274: gap of unknown length
19175     22642: contig of 3368 bp in length
19275     22742: gap of unknown length
22643     22742: gap of unknown length
22743     25848: contig of 3106 bp in length
25848     25948: gap of unknown length
25849     28042: contig of 2094 bp in length
25949     28043: gap of unknown length
28043     28143: contig of 1857 bp in length
28143     30099: gap of unknown length
30099     30100: gap of unknown length
30100     33397: contig of 3298 bp in length
33397     33398: gap of unknown length
33398     33497: gap of unknown length
33497     36954: contig of 3456 bp in length
36954     37053: gap of unknown length
37053     39928: contig of 2875 bp in length
39928     40028: gap of unknown length
40028     42136: contig of 2108 bp in length
42136     42236: gap of unknown length
42236     45566: contig of 3330 bp in length
45566     45567: gap of unknown length
45567     47153: contig of 1487 bp in length
47153     47253: gap of unknown length
47253     48898: contig of 1645 bp in length
48898     48996: gap of unknown length
48996     50300: contig of 1302 bp in length

```

```

50301 50400: gap of unknown length
50401 52989: contig of 2589 bp in length
52990 53089: gap of unknown length
53090 54605: contig of 1516 bp in length
54606 54705: gap of unknown length
54706 57531: contig of 2826 bp in length
57532 57631: gap of unknown length
57632 60066: contig of 2435 bp in length
60067 60166: gap of unknown length
60167 61817: contig of 1651 bp in length
61818 61917: gap of unknown length
61918 63469: contig of 1552 bp in length
63470 65866: gap of unknown length
65867 65966: gap of unknown length
65967 68814: contig of 2848 bp in length
68815 70324: gap of unknown length
70325 70424: gap of unknown length
70425 72206: contig of 1782 bp in length
72207 72306: gap of unknown length
72307 73955: contig of 1649 bp in length
73956 74055: gap of unknown length
74056 75353: contig of 1298 bp in length
75354 77676: contig of 2223 bp in length
77677 79896: contig of 2120 bp in length
79897 81516: gap of unknown length
81517 81616: gap of unknown length
81617 83235: contig of 1619 bp in length
83236 85103: gap of unknown length
85104 86299: gap of unknown length
86299 86899: gap of unknown length
86900 88102: contig of 1203 bp in length
88103 88202: gap of unknown length
88203 89377: contig of 1175 bp in length
89378 89477: gap of unknown length
89478 90752: contig of 1275 bp in length
90753 90852: gap of unknown length
90853 92238: contig of 1386 bp in length
92239 92339: gap of unknown length
92340 94149: contig of 1811 bp in length
94150 94249: gap of unknown length
94250 96535: contig of 2286 bp in length
96536 96635: gap of unknown length
96636 98476: contig of 1841 bp in length
98477 98576: gap of unknown length
98577 99663: contig of 1087 bp in length
99664 101115: gap of unknown length
101116 101215: contig of 1352 bp in length
101216 102350: gap of unknown length
102351 102630: gap of unknown length
102631 104194: contig of 1564 bp in length
104195 104294: gap of unknown length
104295 105839: contig of 1545 bp in length
105840 105939: gap of unknown length
105940 107145: contig of 1206 bp in length
107146 107245: gap of unknown length
107246 108445: contig of 1200 bp in length
108446 108545: gap of unknown length
108546 110126: contig of 1581 bp in length
110127 110226: gap of unknown length
110227 111821: contig of 1595 bp in length
111822 111921: gap of unknown length
111922 113705: contig of 1784 bp in length
113706 113805: gap of unknown length
113806 115339: contig of 1534 bp in length
115340 115439: gap of unknown length

```

```

Query Match
Best Local Similarity 59.6%; Pred. No. 1.2;
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 92 GGGGCGAGGGGGGGGATGAGCCAGGCGCGAGAGAACTCTGAACCTCCGATG 151
Db 144736 GGGGCGAGGGGGGGGATGAGCCAGGCGCGAGAGAACTCTGAACGAGGCGGGGATG 144795
QY 152 CTTCTAGACCGCTCAGACACTCTGGGCGAGCGTGAGAGATT 195
Db 144796 CCGGCGAGGGGCTAATATAACCTTGCGATAGCGGGCGGGGATT 144839

```

```

RESULT 13
AC096681/c
LOCUS
DEFINITION
Canis familiaris clone RP81-228C19, WORKING DRAFT SEQUENCE, 3
ACCESSION
AC096681
VERSION
AC096681.1 GI:15718555
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
dog.
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 145617)
Akhter,N., Ayele,K., Beckstrom-Stenbery,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Grant,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Ido,J.R., Karlins,E., Lario,P., Lee-Lin,S.-O.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Mastello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Stattinop,S., Thomas,J.W., Thomas,P.J., Touchman,D.W.,
Tsurgouk,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 145617)
Green,E.D.
Direct Submission
Submitted (21-SEP-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nigr.nih.gov
----- Project Information
Center project name: cti
Center clone name: 228C19
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Strategy: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 144648 bases at least Q40
Consensus quality: 144809 bases at least Q30
Consensus quality: 144939 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 145417; sum-of-ctgifs
Quality coverage: 10.21x in Q20 bases; agarose-fp
Quality coverage: 10.18x in Q20 bases; sum-of-ctgifs

```

* NOTE: This is a 'working draft' sequence. It currently

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
```

[illegible]

	Matches	67:	Conservative	0:	Mismatches	52:	Indels	0:	Gaps	0:
OY	37	CCGACAAAGCAGCTACTTGGGAAGGTCTGGAAAAATCTCCCTGTTTGCGGGG	96							
Db	58914	CCCACAATCTCTTCGCCAGGCCCCAGGGCAGATAGGCACAAAGACTTGGCTACTTCTGTAGGCG	58973							
OY	97	CAGGCGCGGGGATGAGCCAGGCGCAGAGAGAATCTGGAACATCCGAGATTGCTCT	155							
Db	58974	CTGGGAGATGGGATATGTGCAGGTGGGTCAAGGTTCCCATGGTTGCACCCTTCTTGCTCT	59032							
RESULT 15										
AC084337		176831 bp	DNA	linear	HTG-10-FEB-2002					
LOCUS		Homo sapiens chromosome 11 clone CTD-2010I16 map 11, *** SEQUENCING								
DEFINITION		IN PROGRESS ***, 2 unordered pieces.								
ACCESSION		AC084337								
VERSION		AC084337.4								
KEYWORDS		HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.								
SOURCE		human.								
ORGANISM		Homo sapiens								
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS		Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.								
TITLE		1 (bases 1 to 176831)								
JOURNAL		Birren,B., Linton,L., Nusbaum,C. and Lander,E.								
REFERENCE		Homo sapiens chromosome 11, clone CTD-2010I16								
UNPUBLISHED		2 (bases 1 to 176831)								
AUTHORS		Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,								
		Anderson,S., Batra,N., Bastien,V., Bedalov,F., Boguslavskiy,L.,								
		Bookhalter,B., Brown,A., Burkett,G., Campilongo,A., Castile,A.,								
		Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,								
		DeArillano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,								
		Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,								
		Glenn,L., Grand-pierre,N., Hagos,B., Heathford,A., Horton,L.,								
		Iliev,I., Johnson,R., Jones,C., Kamp,L., Karatas,A., LaRoque,K.,								
		Lamazares,R., Landers,T., Lehoczeky,J., Levine,R., Lieu,C., Liu,G.,								
		McDonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,								
		McPheters,R., Meldrim,J., Menous,L., Mihova,T., Mlenga,V.,								
		Morow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,								
		O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,								
		Pierre,N., Pisanic,C., Pollard,V., Raymond,C., Riback,M., Riley,R.,								
		Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,								
		Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,								
		Strauss,N., Sudramanian,A., Talamas,J., Testaye,S., Theodore,J.,								
		Tirrell,A., Travers,M., Trigilio,J., Vassiliou,H., Viel,R., Vo,A.,								
		Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,								
		Zimmer,A. and Zody,M.								
		Direct Submission								
JOURNAL		Submitted (25-OCT-2000) Whitehead Institute/MIT Center for Genome								
		Research, 320 Charles Street, Cambridge, MA 02141, USA								
		On Feb 10, 2002 this sequence version replaced g1:16974143.								
COMMENT		All repeats were identified using RepeatMasker:								
		Smith, A.F.A. & Green, P. (1996-1997)								
		http://ftp.genome.washington.edu/RM/RepeatMasker.html								
		-- Genome Center								
		Center: Whitehead Institute/ MIT Center for Genome Research								
		Center code: WIBR								
		Web site: http://www-seq.wi.mit.edu								
		Contact: sequence_submission@genome.wi.mit.edu								
		----- Project Information								
		Center project name: LI0956								
		Center clone name: 2010_I_16								

		* NOTE: This is a 'working draft' sequence. It currently								
		* consists of 2 contigs. The true order of the pieces								
		* is not known and their order in this sequence record is								

FEATURES	*	22121	22220:	gap of	100	bp
SOURCE		22221	176831:	contig of	154611	bp in length.
				Location/Qualifiers		
				1..176831		
				/organism="Homo sapiens"		
				/db_xref="taxon:9606"		
				/chromosome="11"		
				/map="11"		
				/clone="CTD-2010116"		
BASE COUNT	46586	a	42824	c	36611	g 47521 t 289 others
ORIGIN						
Query Match,		17.0%:	Score 35.8;	DB 2;	Length	176831;
Best Local Similarity		56.3%:	Pred. No. 2.4;			
Matches 67;	Conservative	0;	Mismatches 52;	Indels 0;	Gaps 0;	
QY	37	CCCGAAGGACACAGCTGACTTAGGGAAGGAGTCTGGGAAANAATTCCTCTGTTTGGGGGGG	96			
DB 101639		CCCGACAAATCTTCCAGGCCCGAGGCGACAATAGGCAAGAGACTTGGCTTCTGTAGGGCC	101698			
QY	97	CAGGGCGGGGAGTGAAGCGAGGCGGAGAGAGACTGTGGAAGACTCCCGAGATGCTCT	155			
DB 101699		CTGGGAGATGGGGATATGGCAGAGTGGGTGACAGGTCCTCCATAGGCTTCACTTCTTGCTCT	101757			

```
Search completed: November 6, 2002, 17:45:48
Job time : 1042.58 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:05:40 ; Search time 7.97326 Seconds
(without alignments)
6469.502 Million cell updates/sec

Title: US-09-833-740-7_COPY_1_210
Perfect score: 210
Sequence: 1 ccgcctgtcttcttcttccccc.....ggattgtgcaaacattcc 210

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCNUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfilest.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	13.8	2674	4	US-09-817-180-1
2	28.4	13.5	8791	5	PCT-US96-01735-5
3	28	13.3	4791	4	US-08-949-155-49
4	27.8	13.2	3228	4	US-09-318-448-20
5	27.8	13.2	4184	2	US-08-785-310A-4
6	27	12.9	1762	6	5185262-2
7	27	12.9	1763	6	5185262-1
8	26.8	12.8	3740	4	US-09-162-274A-6
9	26.6	12.7	7218	1	US-08-233-463-14
10	26.4	12.6	584	4	US-09-328-111-83
11	26.4	12.6	1489	4	US-09-487-445-10
12	26.2	12.5	7210	2	US-08-257-963B-10
13	26.2	12.5	7210	4	US-08-367-841A-10
14	26.2	12.5	7210	5	PCT-US95-07201-10
15	26.2	12.5	22481	4	US-08-367-841A-43
16	26.2	12.5	22481	5	PCT-US95-07201-43
17	26.2	12.5	35060	3	US-08-814-095-7
18	26	12.4	26	3	US-08-845-546-23
19	26	12.4	1392	4	US-09-006-353A-1
20	26	12.4	1933	4	US-08-974-380-1
21	26	12.4	1968	5	PCT-US95-14024-2
22	26	12.4	2153	4	US-08-577-492-31
23	26	12.4	2153	4	US-09-079-630-31
24	26	12.4	6407	2	US-08-616-844-7
25	26	12.4	6407	2	US-08-599-654-7
26	26	12.4	6407	2	US-08-944-868A-7
27	26	12.4	6407	3	US-08-944-423A-7

28	26	12.4	6407	3	US-08-944-496-7	Sequence 7, Appl
29	26	12.4	8106	4	US-09-135-241-1	Sequence 1, Appl
30	25.8	12.3	972	1	US-07-940-605A-9	Sequence 9, Appl
31	25.8	12.3	972	2	US-08-690-096-9	Sequence 9, Appl
32	25.6	12.2	2301	4	US-09-085-199B-8	Sequence 8, Appl
33	25.6	12.2	548	1	US-08-469-667-6	Sequence 6, Appl
34	25.6	12.2	548	4	US-09-224-110-6	Sequence 6, Appl
35	25.6	12.2	548	5	PCT-US95-07289-6	Sequence 6, Appl
36	25.6	12.2	1067	4	US-09-045-193-1	Sequence 1, Appl
37	25.6	12.2	2517	1	US-07-906-930E-1	Sequence 1, Appl
38	25.6	12.2	2939	1	US-07-906-930E-3	Sequence 3, Appl
39	25.6	12.2	15297	4	US-09-817-180-3	Sequence 3, Appl
40	25.6	12.2	51952	3	US-08-947-823-1	Sequence 1, Appl
41	25.4	12.1	434	1	US-08-480-784-34	Sequence 34, Appl
42	25.4	12.1	434	1	US-08-483-553-34	Sequence 34, Appl
43	25.4	12.1	434	1	US-08-487-002-34	Sequence 34, Appl
44	25.4	12.1	434	1	US-08-483-554B-34	Sequence 34, Appl
45	25.4	12.1	434	1	US-08-488-011B-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-817-180-1
; Sequence 1, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:
; APPLICANT: GAN, weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO01183
; CURRENT APPLICATION NUMBER: US/09/817,180
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2674
; TYPE: DNA
; ORGANISM: Human
US-09-817-180-1

Query Match 13.8%; Score 29; DB 4; Length 2674;
Best Local Similarity 54.1%; Pred. No. 2.2;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 55 CTTAGGGAAGTCTCTGGGAAATCTCCCTGTTTGGGGGCGAGGGGGGATGAC 114
DB 1083 CTCGGGAATGAAGAGGACACCCCGGAGCGGGGTGAGCTGTGGCAGAGG 1142
QY 115 CAGGGCGGAGGAAGTCTGGAAGTCCGTAGATTGCTGAGCCGCC 163
DB 1143 CAGTCTCTCAGAGAGCACTGCAGGGGCTGCAGTAGCGCTGTGACGCC 1191

RESULT 2
PCT-US96-01735-5
; Sequence 5, Application PC/TUS9601735
; GENERAL INFORMATION:
; APPLICANT: Marks, Andrew R.
; TITLE OF INVENTION: HUMAN T CELL INOSITOL 1,4,5,-TRISPHOSPHATE
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

```
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01735
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/386,039
FILING DATE: 09-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: KOLE, Lisa B
REGISTRATION NUMBER: 35,225
REFERENCE/DOCKET NUMBER: A30042 - 165/30555
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2628
TELEFAX: 212-765-2519
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
PCT-US96-01735-5

Query Match
Best Local Similarity 13.5%; Score 28.4; DB 5; Length 8791;
Matches 71; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

OY 30 AGGAGTCCGAGAACACAGCTACTTAGGAGAGCTGGGAAAATCTCCCTGCTTT 89
DB 5431 AGCAGACCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5490
OY 90 GGGGGGCGAGGGGGGGGATGACCGAGGCGGAGAGAACTCTGAAGACTCCGCTAGT 149
DB 5491 GGCTGAGATGAGTGTGCGCGAGGTTGACCTTGACCTTGACAGAGAGGGGCTTCCAACT 5550
OY 150 TGTCTAGACCGCTCTGACAGAC 171
DB 5551 AGTTATGACCTCATCATGAAC 5572

RESULT 3
US-08-949-155-49/c
Sequence 49, Application US/08949155
Patent No. 6271436
GENERAL INFORMATION:
APPLICANT: Pledrahita, Jorge A
APPLICANT: Bazer, Fuller W
TITLE OF INVENTION: Compositions and Methods for the
TITLE OF INVENTION: Generation of Transgenic Animal Species
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE AND DURKEE
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: US
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/949,155
```

```
FILING DATE: Concurrently Herewith
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/027,338
FILING DATE: 11-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,094
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W
REGISTRATION NUMBER: 41,071
REFERENCE/DOCKET NUMBER: TAMK-177
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-949-155-49

Query Match
Best Local Similarity 13.3%; Score 28; DB 4; Length 4791;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 73 AATATCTCCCTCTTTGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
DB 193 AGATTCTGTGTCTTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 134
OY 133 CTGAGACTCCGTAGATTGCTTAGACCGCTCAGACT 172
DB 133 GTGAGACTAGGAGAAATGAGCTTGACATCCCTCCATT 94

RESULT 4
US-09-318-448-20/c
Sequence 20, Application US/09318448
Patent No. 6210950
GENERAL INFORMATION:
APPLICANT: Johnson, William G.
APPLICANT: Steenroos, Edward S.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318,448
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
TYPE: DNA
LENGTH: 3228
ORGANISM: Homo sapiens
US-09-318-448-20

Query Match
Best Local Similarity 13.2%; Score 27.8; DB 4; Length 3228;
Matches 56; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 89 TGGGGGCGAGGGGGGGGATGAGCCAGGCGGAGAGAACTCTGAAGACTCCGTAGA 148
DB 1689 TGCTGTGAAGGGGGGTCAGAGAGGCCGACAGAGAGGAGGAGGAGGAGGAGGAGGAG 1530
OY 149 TTGCTAGACCGCTCTGACAGACTCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 191
DB 1629 TGTGCTGACCTCCTGCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1587

RESULT 5
US-08-785-310A-4/c
Sequence 4, Application US/08785310A
Patent No. 5840532
```

Dd	153	CACGCGCAGTGGATTGGCGGGAATCCGTGTAAGAAGCTGGGCACGCCGCAATTGGACTTTGAAGT	212
Oy	140	CTCGTGATGATGCTCTAGACCCTTCACAGACACTCTCGGCGCAGCGTGGAGAAGAT	194
Dd	213	CTCGATAGCTATTCTCTCGACAGCTCGAGCGCATACCTCCTTGTCTAGCTGATGAT	267
 RESULT 7 5185262-1:			
; Patent No. 5185262			
; APPLICANT: KOHAMA, KEIKO; KOBAYASHI, MIKI; KURUSU, YASUROU;			
; YUKAWA, HIDEAKI; FUKUSHIMA, MAKIKO			
; TITLE OF INVENTION: DNA FRAGMENT CONTAINING GENE WHICH			
; ENCODES THE FUNCTION OF STABILIZING PLASMID IN HOST MICROORGANISM			
; NUMBER OF SEQUENCES: 2			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/473,396			
; FILING DATE: 01-FEB-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 385,414			
; FILING DATE: 26-JUL-1989			
; SEQ ID NO:1:			
; LENGTH: 1763			
; 5185262-1			
Oy	Query Match	12.9%; Score 27; DB 6; length 1763;	
Dd	Best local Similarity	52.2%; Pred. No. 8.9;	
	Matches 60; Conservative	0; Mismatches 55; Indels 0; Gaps 0	
Oy	80	CCCTGCTTTTGGGGGGGCGAAGGCGGGGGGATGAGCCAGAGGCGCAGAAAGNACTGAGA	139
Dd	154	CACCTGCAGTAGATTGGCGGGAATCCGCTGAAGAAGCTGGAGCGCACFTGGACTTTGAAGT	213
Oy	140	CTCGGTAGATTGCTCTAGACCGCCTCAGACACTCTCGGCGCAGCGTGGAGAAGAT	194
Dd	214	CTCGATAGCTATTCTCTGACAGCTCGAGCGCATACCTCCTTGTCTAGCTGATGAT	268

```

RESULT 8
US-09-162-274A-6/c
; Sequence 6, Application US/09162274A
; Patent No. 6316188
; GENERAL INFORMATION:
; APPLICANT: Yan, Lan
; APPLICANT: Liggett, Stephen B.
; APPLICANT: Galinsky, Raymond E.
; APPLICANT: Weinsliboun, Richard M.
; APPLICANT: Lachman, Herb M.
; TITLE OF INVENTION: ' HISTAMINE-N-METHYLTRANSFERASE VARIANTS
; FILE REFERENCE: 07039/080001
; CURRENT APPLICATION NUMBER: US/09/162,274A
; CURRENT FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3740
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(3740)
; OTHER INFORMATION: n = A,T,C or G
US-09-162-274A-6

Query Match      12.8%; Score 26.8; DB 4; Length 3740;
Best Local Similarity 53.9%; Pred. No. 14;
Matches 55; Conservative 0; Indels 47; Gaps 0;

```

[illegible]

OY 84 GCTTTGGGGGGGCGGCGGCGGATGACCGCCGAG 125
DB 3227 GGTTCGAGAGATTAGCGGTGGGAGTGAATGAGGCCGAG 3186

RESULT 9

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOMLPX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PT99PL-F15
US-08-232-463-14

Query Match 12.7% Score 26.6; DB 1; Length 7218;
Best Local Similarity 11.4% Pred. No. 21;
Matches 17; Conservative 74; Mismatches 58; Indels 0; Gaps 0;

OY 29 AAGGAGCTCCGAGAAAGCAGCTGACTTAGGAGAGCTGGGAAATCTCGCTTT 88
DB 1181 RRR 1122
OY 89 TGGGGGGGCGGCGGCGGATGACCGAGGCGGCGGAGGAAAGCTGGAAGCTCGTAGA 148
DB 1121 RRR 148
OY 149 TTGCTCTAGACCGCTCAGACACTCTCGG 177
DB 1061 AAGCTCCTCGACTGCGAGCAAGCTCGG 1033

RESULT 10
US-09-328-111-83

Sequence 83, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll, III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 83
LENGTH: 584
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(584)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-83

Query Match 12.6% Score 26.4; DB 4; Length 584;
Best Local Similarity 57.1% Pred. No. 9.2;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 115 CAGGGCGGAGGAGAAAGCTGGAAGACTCGTAGATTGCTCTAGACCGCTCAGACACTCT 174
DB 426 CAGGTTCCAGACTGACCGCCCAAGATCCGAGCGTTCTGGGCCACCTTCAAGTGAAC 485
OY 175 CGGCGGCGGCTGAGAGATTGT 198
DB 486 GGGGCAACATGATGCTTGT 509

RESULT 11
US-09-487-445-10
Sequence 10, Application US/09487445
Patent No. 6258600
GENERAL INFORMATION:
APPLICANT: Hong Zhang
APPLICANT: Lex M. Cowart
TITLE OF INVENTION: ANTISENSE MODULATION OF CASPASE 8 EXPRESSION
FILE REFERENCE: RTS-0107
CURRENT APPLICATION NUMBER: US/09/487,445
CURRENT FILING DATE: 2000-01-19
NUMBER OF SEQ ID NOS: 176
SEQ ID NO 10
LENGTH: 1489
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (47)...(1489)
US-09-487-445-10

Query Match 12.6% Score 26.4; DB 4; Length 1489;
Best Local Similarity 55.4% Pred. No. 13;
Matches 51; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 5 CTGTCTCTTCTCCTCAGCCTGCAAGAGATCCGAGAAAGCAGCTGACTTAGGAAG 64

Db 1114 CTTTTCATTTCAGGCTTGCAGAGAACTTAATCTCAGAAAGAGTGCCTGATGAGGCAGG 1173
QY 65 GTCTGGGAAAATCTCCCTGTTTGGGGGG 96
Db 1174 CTTGAGACACAGAACCACTTTAGAAGTGG 1205

RESULT 12

US-08-257-963B-10
; Sequence 10, Application US/08257963B
; Patent No. 5840866
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, S.
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Tanikaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Flinnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257, 963B
; FILING DATE:
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEPT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AOTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7210 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; IMMEDIATE SOURCE:
; LIBRARY: DASH II
; FEATURE:
; NAME/KEY: JT106
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 7.2 kb No. 5840866 1 fragments
; OTHER INFORMATION: Derived from human placental genomic DNA
US-08-257-963B-10

Query Match 12.5%; Score 26.2; DB 2; Length 7210;
Best Local Similarity 53.4%; Pred. No. 28;
Matches 55; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 22 GCCTGTCAAGAGTCCAGAAACACAGCTGACTTGAAGGAAAGTCTGGGAAAAATCTCC 81
Db 6869 GGCACCTCAGAGGACAGAAAAAGAGGGGTGCAAGGAGAGAAATGCGAGACAGAGCC 6928

QY 82 CTGCTTTTGGGGGCGAGGGCGGGGATGAGCCAGGCCGAG 124
Db 6929 CTCGAAATTTGGCGCAAAAGGTGAGTGATGAGAGAGGCGAGAG 6971

RESULT 13

US-08-367-841A-10
; Sequence 10, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tombran-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Flinnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367, 841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/257, 963
; FILING DATE: 07-JUN-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AOTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7210 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; IMMEDIATE SOURCE:
; LIBRARY: DASH II
; FEATURE:
; NAME/KEY: JT6A
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 7.0 kb No. 6319687 1-No. 6319687
; OTHER INFORMATION: fragment; Derived from human placental
; OTHER INFORMATION: genomic DNA; also referred to as JT106
US-08-367-841A-10

Query Match 12.5%; Score 26.2; DB 4; Length 7210;
Best Local Similarity 53.4%; Pred. No. 28;
Matches 55; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 22 GCCTGTCAAGAGTCCAGAAACACAGCTGACTTGAAGGAAAGTCTGGGAAAAATCTCC 81
Db 6869 GGCACCTCAGAGGACAGAAAAAGAGGGGTGCAAGGAGAGAAATGCGAGACAGAGCC 6928

Qy 82 CTCGCTTTTGGGGGGGAGGGGCGGATGAGCCAGGGCCGAG 124
 ||| ||| ||| | ||||| ||||| |||
 Db 6921 CTGCAATTTGGGCAAAAGCGTGCATGAGAGAGGGCGCAGAG 6963

```
Search completed: November 6, 2002, 14:20:31
Job time : 42.9733 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:17:21 ; Search time 260.086 Seconds
(without alignments)
10897.801 Million cell updates/sec

Title: US-09-833-740-7_COPY_1_210

Perfect score: 210
Sequence: 1 ccgcctgtctctctctc.....ggattgtgaacatttc 210

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estbun:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52.6	25.0	689	12	AZ971837 2M0245B02
2	38.6	18.4	772	9	AU119486 AU119486
3	34.8	16.6	939	10	BG855275 1024042C0
4	34.2	16.3	771	10	BG479234 602526449
5	34	16.2	623	9	AA690752 vus7c08.r
6	33.8	16.1	895	10	BG755847 602716308
7	33.6	16.0	695	10	BI256085 602978917
8	33.6	16.0	836	12	AG076601 Pan trogl
9	33.2	15.8	258	10	BI831409 603074508
10	33.2	15.8	303	10	BF172864 MYE0121 M
11	33.2	15.8	545	12	CNS0172C AL167169 Tetradon
12	33.2	15.8	597	10	BE336705 ba99d01.Y
13	33.2	15.8	662	10	BI828928 603075090
14	33.2	15.8	882	10	BG035198 60324706
15	33	15.7	386	10	BF759824 PMO-CT064
16	32.8	15.6	921	12	AQ898422 HS_315_B
17	32.8	15.6	1101	12	CNS017R9 AL108399 Drosoph11

C	18	32.6	15.5	430	9	BB776545	BB776545
	19	32.6	15.5	490	9	AI561790	AI561790 vv6b12.x
	20	32.6	15.5	1019	12	CNS05MS	AL43657 Tetradon
	21	32.4	15.4	288	9	AA788598	AA788598 ag58q10.s
	22	32.4	15.4	300	9	AW517011	AW517011 xp90c11.x
	23	32.4	15.4	334	10	BF001888	BF001888 7994h12.x
	24	32.4	15.4	356	10	BF195077	BF195077 7094f04.x
	25	32.4	15.4	377	9	AA776500	AA776500 ah11f0.s
	26	32.4	15.4	380	9	AA439335	AA439335 xc17g05.x
	27	32.4	15.4	407	9	AI874207	AI874207 km50c02.x
	28	32.4	15.4	411	9	AA473156	AA473156 xy14h02.x
	29	32.4	15.4	412	10	BM127741	BM127741 1f08c08.x
	30	32.4	15.4	413	9	AA769135	AA769135 xj19a09.x
C	31	32.4	15.4	428	10	BM127990	BM127990 1f08c08.y
	32	32.4	15.4	447	9	AI804686	AI804686 lt94c04.x
	33	32.4	15.4	461	9	AI184928	AI184928 ok28c02.x
	34	32.4	15.4	464	9	AI201132	AI201132 qf64f09.x
	35	32.4	15.4	502	9	AA779225	AA779225 zj39e07.s
	36	32.4	15.4	567	9	AA543126	AA543126 vk36b12.r
	37	32.4	15.4	574	10	BI438774	BI438774 1c26e06.x
C	38	32.4	15.4	674	12	AG002186	AG002186 Homo sapi
	39	32.4	15.4	687	12	AG002187	AG002187 Homo sapi
C	40	32.4	15.4	707	12	AG002182	AG002182 Homo sapi
	41	32.4	15.4	750	12	AG002170	AG002170 Homo sapi
	42	32.4	15.4	755	12	AG002169	AG002169 Homo sapi
	43	32.4	15.4	777	10	BI463395	BI463395 60320431
C	44	32.2	15.3	305	9	AA153765	AA153765 mc04b03.r
	45	32.2	15.3	519	9	AU175395	AU175395 AU175395

ALIGNMENTS

RESULT 1
LOCUS AZ971837/c 689 bp DNA linear GSS 27-Apr-2001
DEFINITION 2M0245B02R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0245B02 R, DNA sequence.
ACCESSION AZ971837
VERSION AZ971837.1 GI:1843064
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 689)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0245 row: B column: 02
Seq primer: CACACAGGAACACGATATGACC
Class: Plasmid ends
High quality sequence stop: 689.
Location/Qualifiers
1..689
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0245B02"
/clone_1lb="Mouse 10kb plasmid UUGC2M library"

/sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD22ny; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD22 ([g11473211419b1/fl12072.1](#)), a copy-number
 inducible derivative of pMS101 R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT ORIGIN	173 a	189 c	147 g	180 t
----------------------	-------	-------	-------	-------

Query Match	25.0%;	Score 52.6;	DB 12;	Length 689;
Best Local Similarity	68.5%;	Pred. No. 0.00013;		
Matches 87;	Conservative 0;	Mismatches 39;	Indels 1;	Gaps 1;

```

0y      3  GCCCTGTCCTTCTCTCCACGCCCTTCAGCAAGTCCCGAAGACACAGCTGACTT-AGGG  61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
db     137  GTCCTGCTTTTCTCTCGGGCTCTGAGCAAGTCCCGAGGACGCTAGACGCTTGGGGG  78

```

```

0y      62 AAGTCTGGGAAAAATCTCCCTGCTTTTGCGGGGCAGGGGCCGGGGGATTAGACCAGGGCC 121
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      77 TAGTCTGGGAAAATCTCCCAAGATTTTAGAGGGGCACGGCGGGCATGAACAATGGA 18

```

QY	122	GAGAGG	128
Db	17	GATTCG	11

RESULT	2
AU119486/c	
LOCUS	AU119486 HEMBA1 Homo sapiens CDNA clone HEMBA100531 5' mRNA
DEFINITION	sequence.
ACCESSION	AU119486
VERSION	AU119486.1 GI:10934721
KEYWORDS	EST.
SOURCE	human.

REFERENCE
1 (bases 1 to 772)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE	HRI human cDNA project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Takao Isogai

Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952

Email: genomics@hri.co.jp
HRI human CDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute. cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES	Location/Qualifiers
source	1. .772
	/organism="Homo sapiens"

```

/db_xref="taxon:6060"
/clone="HEMBA1005931"
/clone_1lb="HEMBA1"
/tissue-type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/notes="vector: pME18SFL"

```

Query Match	18.4%	Score	38.6	DB	9	Length	772
Best Local Similarity	52.1%	Pred	0.13				
Matches	86	Conservative	0	Mismatches	79	Indels	0
						Gaps	0

37 CCAGGAAGCACAGCTGACTTAGGAGAGCTCTGGAAAAATCTCCCTGTTTTGGGGGG 96
 258 CCCAAGAGCACAGGAGCCACCTGAGGAAAGACAGGAGGACAGCTCTCCAGCTGGGCACCT 199

QY 97 CAGGGCCGGGGGATGAGCCAGGGCCGAGAAAGAACTCTGAGACCTCGTAGATTGCTCTA 156
 198 TAAATCTGAGAGAGAAAGCTCTTGCTCTGGGGAACATTACAGAGCCTGGACCATAGGTTA 139
 Db

Qy 157 GACCGCCCTCAGACACTCTCGGGCGAGCGTGAGAGGATTTGTGCA 201
 ||| | | | | | | | | | | | | | | |
 Db 138 CACCTCATAGGGAAGACAGCGCTGATCCTGTTGAGGGTTTGCCA 94

RESULT 3
B0855275/C
DOCECATE
000 1-
END
PAGE 00 NEW 0001

DEFINITION	1024042C07.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
ACCESSION	BG855275
VERSION	BC085527.1 CT.14236A50

KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota: Viridiplantae: Chlorophyta: Chlorophyceae: Volvocales.

REFERENCE
AUTHORS

Chlamydomonadaceae: Chlamydomonas.
1 (bases 1 to 939)
Grossman, A., Davies, J., Federpiel, N., Harris, E., Lefebvre, P.,
McDermott, J. P., Silflow, C., Stern, D. and Surzwick, R.

TITLE
Analyses of the *Chlamydomonas reinhardtii* Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
Unpublished (2000)

COMMENT:
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000

FEATURES

Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

Location/Qualifiers

```

source
1. .339
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db xref="taxon:1055"

```

In
 /note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI: This library, constructed by John Davies and Jeffery
 McDermott combines cDNAs from CC-1690 cells grown to

light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO₂ and HS medium bubbled with 5% CO₂. PolyA mRNA was purified from each sample, pooled and cDNA synthesized using diethylamyl ethylcarbazate for 10 cycles.

plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described

BASE COUNT	199 a	336 c	140 g	264 t
------------	-------	-------	-------	-------

Db 228 GCGTAAGCCAGACGAGGCGTAGCATTTCTCAAGATTTCAGACTGCTGCTAC 175

RESULT 6

LOCUS
Bg755847

DEFINITION 895 bp mRNA linear EST 15-MAY-2001

ACCESSION 602716308F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:485665 5'

VERSION Bg755847

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 895)

AUTHORS NIH-MGC http://mhc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM1707 row: 0 column: 02

High quality sequence stop: 816.

Location/Qualifiers

1. .895

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:485665"

/tissue="Primary B-cells from tonsils (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pORF7; Site: 1: XhoI;

Site: 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGG(6). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: This is a NIH-MGC Library."

BASE COUNT 177 a 284 c 292 g 141 t

ORIGIN

Query Match

Best Local Similarity 56.4%; Score 33.8; DB 10; Length 895;

Matches 62; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11302 row: b column: 06

High quality sequence start: 5

High quality sequence stop: 679.

Location/Qualifiers

1. .695

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5123621"

/tissue="Muscle"

/lab_host="N1 CGAP L19"

/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 148 a 223 c 187 g 137 t

ORIGIN

Query Match

Best Local Similarity 53.9%; Score 33.6; DB 10; Length 695;

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11302 row: b column: 06

High quality sequence start: 5

High quality sequence stop: 679.

Location/Qualifiers

1. .695

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5123621"

/tissue="Muscle"

/lab_host="N1 CGAP L19"

/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 148 a 223 c 187 g 137 t

ORIGIN

Query Match

Best Local Similarity 53.9%; Score 33.6; DB 10; Length 695;

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11302 row: b column: 06

High quality sequence start: 5

High quality sequence stop: 679.

Location/Qualifiers

1. .695

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5123621"

/tissue="Muscle"

/lab_host="N1 CGAP L19"

/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 148 a 223 c 187 g 137 t

ORIGIN

Query Match

Best Local Similarity 53.9%; Score 33.6; DB 10; Length 695;

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11302 row: b column: 06

High quality sequence start: 5

High quality sequence stop: 679.

Location/Qualifiers

1. .695

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5123621"

/tissue="Muscle"

/lab_host="N1 CGAP L19"

/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 148 a 223 c 187 g 137 t

ORIGIN

Query Match

Best Local Similarity 53.9%; Score 33.6; DB 10; Length 695;

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11302 row: b column: 06

High quality sequence start: 5

High quality sequence stop: 679.

Location/Qualifiers

1. .695

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5123621"

/tissue="Muscle"

/lab_host="N1 CGAP L19"

/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 148 a 223 c 187 g 137 t

ORIGIN

Query Match

Best Local Similarity 53.9%; Score 33.6; DB 10; Length 695;

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11302 row: b column: 06

High quality sequence start: 5

High quality sequence stop: 679.

Location/Qualifiers

1. .695

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5123621"

/tissue="Muscle"

/lab_host="N1 CGAP L19"

/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 148 a 223 c 187 g 137 t

ORIGIN

Query Match

Best Local Similarity 53.9%; Score 33.6; DB 10; Length 695;

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

LOCUS

DEFINITION

ACCESSION

VERSION

COMMENT

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbesc@gscc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

FEATURES

source

Sequencing: -21M13
LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .836

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-070N17.F"

/sex="male"

/cell_type="lymphoblast"

/clone_1ib="PTB Chimpanzee Male BAC Library"

BASE COUNT 145 a 178 c 237 g 261 t 15 others

ORIGIN

Query Match 16.0%; Score 33.6; DB 12; Length 836;

Best Local Similarity 51.3%; Pred. No. 35;

Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 48 CAGCTGACTTAGGAGGAGCTGCGGAAATCTCCCTTTTGGGGGCGGCGGCGG 107

DB 253 CGGCAAGTCTTGGTTGGAGCCCAATCTTCCCGCAGAGGGGCGGCGGCGTGC 312

QY 108 GATGAGCCAGGGCCGAGAGAACTCTGAACTCCCTGATGCTCTGACCGCTCAG 167

DB 313 GGTTCGCGGGGGGAGAGCAATTAAGTCTGCTTGTGTGTTTAAACCGCAAT 372

QY 168 ACATCTGGCGGCGAGCGAGAGATTTGTG 199

DB 373 TAACTACTACCCAGTGTGCTGGGCTGTG 404

RESULT 9 258 bp mRNA linear EST 04-OCT-2001

BI831409/c LOCUS 603074508F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:516577 5'

DEFINITION mRNA sequence.

ACCESSION BI831409

VERSION BI831409.1 GI:15942959

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 258)

NIH-MGC http://mgs.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9apds-femail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: L14M1413 row: P column: 02

High quality sequence stop: 258.

location/Qualifiers

1. .258

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:516577"

BASE COUNT

45 a 97 c 69 g 47 t

ORIGIN

Query Match 15.8%; Score 33.2; DB 10; Length 258;

Best Local Similarity 67.1%; Pred. No. 31;

Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 59 GCGAAGTCTGGGAAATCTCCCTTTTGGGGGCGGCGGCGGATGAGCCAGG 118

DB 209 GGGAGGACGTGGGCAATGCCCCCTGCTTGGGTGGAATAGTGTGGTAAAGCAGG 150

QY 119 GCCGAGAAG 128

DB 149 GCCGAGAAG 140

RESULT 10

BF172864

LOCUS

DEFINITION

sequence.

ACCESSION

BF172864

VERSION

BF172864.1 GI:13439132

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 303)

Claudio J.O., Tang H., Khan E.M., Voralia M., Li Z., Cukerman E.,

Francisco-Pabalan O., Ilew C.C. and Stewart A.K.

The transcriptional phenotype of myeloma cells

Unpublished (2000)

Contact: A. Keith Stewart, M.D.

Oncology Research

University Health Network

610 University Ave., 5-126, Toronto, Ontario, M5G 2M9, Canada

Tel: (416) 946-4639

Fax: (416) 946-6546

Email: k.stewart@utoronto.ca

PCR Primers

FORWARD: 5'-GCCAAGCTCGAATTAACCTCACTAAGG-3'

BACKWARD: 5'-CGAATGATGTATAGACCTCACTAAGGCG-3'

Seq primer: 5'-GAATTAACCTCACTAAGG-3'

FEATURES

source

location/Qualifiers

1. .303

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_1ib="Myeloma (MYE) cDNA library"

/sex="male"

/tissue_type="Blood"

/cell_type="myeloma"

/dev_stage="multiple myeloma"

/note="Vector: Lambda Zap Express; site_1: EcoRI; site_2: XhoI; Myeloma cells from multiple myeloma patients' bone marrow were purified by magnetic cell sorting. mRNA was purified and an oligo d(T)18 primer containing XhoI restriction site was used to prime first strand synthesis using M-MV reverse transcriptase. To protect the cDNAs from XhoI digestion in subsequent cloning step, the nucleotide analogue 5-methyl-dCTP was added to the

RESULT 13
BI828928 662 bp mRNA linear EST 04-OCT-2001
LOCUS 603075090F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:516694 5',
DEFINITION mRNA sequence.
ACCESSION BI828928
VERSION BI828928.1 GI:15940478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 662)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11414 row: d column: 23
High quality sequence stop: 662.
Location/Qualifiers
1. 662
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:516694"
/clone_1lb="NIH_MGC_119"
/issue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."
BASE COUNT 111 a 227 c 203 g 121 t
ORIGIN
Query Match 15.8%; Score 33.2; DB 10; Length 662;
Best Local Similarity 67.1%; Pred. No. 42;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 59 GGAAGCTGTGGGAAATCTCCCTGTTTGGGGGCGGCGGGGATGAGCCAGG 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 613 GGGAGGACCTGGGCAATGGCCCCCTGCTGGGTGGAATAGTGGTGTAAAGGCGAGG 554
QY 119 GCCGGAAGG 128
||||| |||||
Db 553 GCCGGGAGG 544

RESULT 14
BG035198 882 bp mRNA linear EST 24-JAN-2001
LOCUS 603324706F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4413048 5',
DEFINITION mRNA sequence.
ACCESSION BG035198
VERSION BG035198.1 GI:12429091
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10138 row: k column: 01
High quality sequence stop: 690.
Location/Qualifiers
1. 882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4413048"
/clone_1lb="NIH_MGC_90"
/issue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and was constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 197 a 308 c 215 g 162 t
ORIGIN
Query Match 15.8%; Score 33.2; DB 10; Length 882;
Best Local Similarity 67.1%; Pred. No. 46;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 59 GGAAGCTGTGGGAAATCTCCCTGTTTGGGGGCGGCGGGGATGAGCCAGG 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 387 GGGAGGACCTGGGCAATGGCCCCCTGCTGGGTGGAATAGTGGTGTAAAGGCGAGG 328
QY 119 GCCGGAAGG 128
||||| |||||
Db 327 GCCGGGAGG 318

RESULT 15
BF759824 386 bp mRNA linear EST 12-JAN-2001
LOCUS PMO-CT0642-151200-001-g01 CT0642 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF759824
ACCESSION BF759824
VERSION BF759824.1 GI:12107724
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 386)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-PM0&l2-PM0-CT0642-151200-001-901&l3-2000-12-15&l4-1)

Seq primer: puc 18 forward

High quality sequence start: 22

High quality sequence stop: 109.

Location/Qualifiers

FEATURES

source
1..386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0642"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 114 a 91 c 100 g 81 t

ORIGIN

15.7%; Score 33; DB 10; Length 386;

Best Local Similarity 53.5%; Pred. No. 40;

Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
OY 63 AGCTGTGGGAAAATCTCCCTGCTTTGGGGGGCGAGGGGGGATGAGCCAGGCCG 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 AGCTCTTGGAACCTGTGCCAGGCTTGAGGGTGGAGAGGCACTTCTAGAGCCGACCCG 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 123 AGAGGAACCTGTAAGACTCCGTAGATTGCTCTAGACCCGCTCAGACACTCTCGGCCAG 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 CCAAGCCAGCAAGCAATGAGCCCATCTCTCTCGCAGAGCAGAGAAATACCTGGGGGCAG 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 183 CGTGGAGAG 191
    ||| |||
Db 313 AGGCCATAG 321
    ||| |||
```

Search completed: November 6, 2002, 16:07:08
Job time : 269.086 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:05:40 : Search time 63.0267 Seconds
(without alignments)
6469.302 Million cell updates/sec

Title: US-09-833-740-1_COPY_1_1660

Perfect score: 1660

Sequence: 1 aggttaccgcatttgaca.....ggaagagacttgcaaca 1660

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCPNUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	3.5	3066	1 US-08-142-439A-1	Sequence 1, Appli
2	58	3.5	3066	2 US-08-869-477-1	Sequence 1, Appli
3	58	3.5	3381	4 US-09-009-119-1	Sequence 1, Appli
4	58	3.5	3381	4 US-09-371-507-1	Sequence 1, Appli
5	58	3.5	3383	5 PCT-US95-09098-1	Sequence 1, Appli
6	58	3.5	10409	3 US-08-772-440-3	Sequence 3, Appli
7	57.6	3.5	200	1 US-08-222-177A-8	Sequence 8, Appli
8	57.4	3.5	48974	4 US-08-920-422-17	Sequence 17, Appli
9	57.2	3.4	372	3 US-08-750-064-1	Sequence 1, Appli
10	57.2	3.4	372	3 US-08-545-196B-14	Sequence 14, Appli
11	57.2	3.4	80246	4 US-08-222-177A-14	Sequence 4, Appli
12	57	3.4	223	1 US-08-720-464A-1	Sequence 1, Appli
13	56.4	3.4	3854	3 US-08-953-823A-1	Sequence 14, Appli
14	56.4	3.4	7218	1 US-08-232-463-14	Sequence 14, Appli
15	56.4	3.4	818	4 US-08-205-697A-4	Sequence 4, Appli
16	56.2	3.4	818	4 US-08-702-525-4	Sequence 4, Appli
17	56.2	3.4	818	5 PCT-US95-02576-4	Sequence 4, Appli
18	56.2	3.4	1570	4 US-08-205-697A-10	Sequence 10, Appli
19	56.2	3.4	1570	4 US-08-702-525-10	Sequence 10, Appli
20	56.2	3.4	1570	5 PCT-US95-02576-10	Sequence 64, Appli
21	56.2	3.4	1605	4 US-08-702-525-64	Sequence 64, Appli
22	56.2	3.4	1605	5 PCT-US95-02576-64	Sequence 1, Appli
23	56.2	3.4	1888	4 US-08-205-697A-1	Sequence 1, Appli
24	56.2	3.4	1888	4 US-08-702-525-1	Sequence 1, Appli
25	56.2	3.4	1888	5 PCT-US95-02576-1	Sequence 3, Appli
26	56.2	3.4	2516	4 US-08-205-697A-3	Sequence 3, Appli
27	56.2	3.4	2516	4 US-08-205-697A-3	Sequence 3, Appli

28	56.2	3.4	2516	4 US-08-702-525-3	Sequence 3, Appli
29	56.2	3.4	2516	5 PCT-US95-02576-3	Sequence 3, Appli
30	55.6	3.3	1906	4 US-09-031-962D-5	Sequence 5, Appli
31	55.6	3.3	80595	4 US-09-078-294-3	Sequence 3, Appli
32	55.4	3.3	965	3 US-09-280-799-133	Sequence 133, App
33	55.4	3.3	3925	4 US-08-793-044-1	Sequence 1, Appli
34	55.4	3.3	8387	2 US-08-532-814-1	Sequence 1, Appli
35	55.4	3.3	8388	4 US-09-225-509-1	Sequence 1, Appli
36	55.2	3.3	3172	1 US-07-741-940-3	Sequence 3, Appli
37	55.2	3.3	3172	1 US-08-289-548A-3	Sequence 3, Appli
38	55.2	3.3	3172	1 US-08-452-654-3	Sequence 3, Appli
39	55.2	3.3	3172	1 US-08-452-655B-3	Sequence 3, Appli
40	55.2	3.3	3172	3 US-08-450-582-3	Sequence 3, Appli
41	54.8	3.3	4157	2 US-08-871-266B-1	Sequence 1, Appli
42	54.8	3.3	4157	2 US-08-819-458A-1	Sequence 1, Appli
43	54.8	3.3	4157	2 US-09-018-864A-1	Sequence 1, Appli
44	54.8	3.3	4157	3 US-08-871-267B-1	Sequence 1, Appli
45	54.8	3.3	4157	4 US-09-618-419-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-142-439A-1
Sequence 1, Application US/08142439A
Patent No. 5670360
GENERAL INFORMATION:
APPLICANT: Thorens, Bernard
TITLE OF INVENTION: Receptor for the Glucagon-Like Peptide-1
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56703600 No. 5670360disk of No. 5670360th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/142,439A
FILING DATE: 24-NOV-93
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 398/92
FILING DATE: 25-MAR-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/00697
FILING DATE: 23-MAR-93
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 3756.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3066 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Rat
FEATURE:
NAME/KEY: CDS

Db 1772 ACACACACACACACACACACACACACACACACACACATTTC 1713
QY 287 TTCCTCCCTCCACTTCCTCC 308
1712 GTCCTGAGCCCGAACCCCGC 1691

RESULT 5

PCT-US95-09098-1/c

Sequence 1, Application PC/TUS9509098

GENERAL INFORMATION:

APPLICANT: Sato, Ryo

APPLICANT: Boynton, John E.

APPLICANT: Gillham, Nicholas W.

APPLICANT: Harris, Elizabeth H.

TITLE OF INVENTION: Porphyrin-Accumulating Type Herbicide

TITLE OF INVENTION: Resistance Gene

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

STREET: 810 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/09098

FILING DATE: 20-JUL-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 2185-110P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3383 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Chlamydomonas reinhardtii

SPRAIN: RS-3

PCT-US95-09098-1

Query Match

Best Local Similarity 81.7%; Score 58; DB 5; Length 3383;

Matches 67; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 227 AACACACACACACACACACACACACACACACACACACATTTTC 286

Db 1773 ACACACACACACACACACACACACACACACACACACATTTC 1714

QY 287 TTCCTCCCTCCACTTCCTCC 308

Db 1713 GTCCTGAGCCCGAACCCCGC 1692

RESULT 6

US-08-772-440-33

Sequence 33, Application US/08772440

Patent No. 6046158

GENERAL INFORMATION:

APPLICANT: Aritzumi, Kiyoshi

APPLICANT: Takashima, Akira

TITLE OF INVENTION: UNIQUE DENDRITIC CELL-ASSOCIATED C-TYPE

TITLE OF INVENTION: LECTINS, DECTIN-1 AND DECTIN-2; COMPOSITIONS AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/772,440

FILING DATE: CONCURRENTLY HERewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.

REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: UTXD:493

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 10409 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: modified_base

LOCATION: 6510

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "D" = A or G or T"

FEATURE:

NAME/KEY: modified_base

LOCATION: 3406..6470

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "K" = G or T"

FEATURE:

NAME/KEY: modified_base

LOCATION: 3564..7896

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "M" = A or C"

FEATURE:

NAME/KEY: modified_base

LOCATION: 3497..3607

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "N" = A or C or G or T"

FEATURE:

NAME/KEY: modified_base

LOCATION: 3479..6422

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "R" = A or G"

FEATURE:

NAME/KEY: modified_base

LOCATION: 3405..6871

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "S" = C or G"

FEATURE:

NAME/KEY: modified_base

LOCATION: 3457..9998

OTHER INFORMATION: /mod_base= OTHER

OTHER INFORMATION: /note= "W" = A or T"

FEATURE:

Patent NO. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree

RESULT 14
 US-08-953-823A-1
 : Sequence 1, Application US/08953823A
 : Patent No. 6136958
 : GENERAL INFORMATION:
 : APPLICANT: Desauvage, Frederic
 : APPLICANT: Rosenthal, Arnon
 : APPLICANT: Stone, Donna
 : TITLE OF INVENTION: Vertebrate Smoothened Proteins
 : NUMBER OF SEQUENCES: 5
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Genentech, Inc.
 : STREET: 1 DNA Way
 : CITY: South San Francisco
 : STATE: California
 : COUNTRY: USA
 : ZIP: 94080
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: 3.5 Inch, 1.44 Mb floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Winpatin (Genentech)
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/953,823A
 : FILING DATE: 30-Sep-1997
 : CLASSIFICATION: 424
 : PRIORITY APPLICATION DATA:
 : APPLICATION NUMBER: 60/027070

RESULT 15
 US-08-232-463-14/c
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367
 ; GENERAL INFORMATION:
 ; APPLICANT: DORNER, F.
 ; APPLICANT: SCHEIFLINGER, F.
 ; APPLICANT: FALKNER, F. G.
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 ; NUMBER OF SEQUENCES: 52
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 1800 Diagonal Road, Suite 500
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22313-0299
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/232,463
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/935,313
 ; FILING DATE:
 ; APPLICATION NUMBER: EP 91 114 300.6
 ; FILING DATE: 26-AUG-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,766
 ; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703)836-9300
 ; TELEFAX: (703)683-4109
 ; TELEX: 899149
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7218 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single

TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PT99C-F1s
US-08-232-463-14

Query Match 3.4%, Score 56.4; DB 1; Length 7218;
Best Local Similarity 10.3%; Pred. No. 2.3e-07;
Matches 42; Conservative 194; Mismatches 170; Indels 0; Gaps 0;

```

OY 385 TTCGAGTGTATCTTTTATCATTAAGTTGAATCTTCCACGTGTGGTTTAAAGT 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1531 TTCTAGACCATCTATCTACATTCAAAAACGCATGTAGCATCTGTAATTACCTAT 1472
OY 445 TTTAGGATTTTATCTAGCGCATCTCCCTCCCTGTAATGTCAGAACTTCACT 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1471 CTATCAGTACTTAAGAGATAGAAGATTGGTACRRRRRRRRRRRRRRRRRRR 1412
OY 505 GGGCTTGCTACCTAATGGAATGATCTATGTTGACTTAATGTGAAGAAAAA 564
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1411 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352
OY 565 GAAGGGGAAAGAGAGGAGGAGAAAGAGGAAAGAACTGCTTTATGCTATT 624
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1351 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1292
OY 625 GCTACTTAACATTTTGTCTCACCCTTGTCTTCAATGGAAGACTGATAG 684
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1291 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1232
OY 685 AAGCTGGAGCCAGCCAGGATAGAGAGTGTGTGTGGGGGGGTGGGAG 744
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172
OY 745 CAAGCAGACCTTAGACAGAGAGAGCCTGTAGAGATGAG 790
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1126

```

Search completed: November 6, 2002, 14:19:56
Job time : 417.027 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:03:45 ; Search time 259.209 Seconds
(without alignments)
10995.308 Million cell updates/sec

Title: US-09-833-740-1_COPY_1_1660

Perfect score: 1660

Sequence: 1 aggttaccgacttgcaca.....gaagagacttgccaaca 1660

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802.*
1: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1658.8	99.9	2170	24	AAH43758
2	64.6	3.9	1199	14	AAQ36051
3	63.4	3.8	245	19	AAV30918
4	63.4	3.8	245	22	AAF98397
5	61.8	3.7	353	22	AAI88828
6	61.2	3.7	3809	12	AAO12760
7	60.6	3.7	6183	24	ABL33197
8	60.6	3.7	23885	22	AAK70103
9	60.6	3.7	23885	22	AAK73173

10	60.4	3.6	415	22	AAI90376
11	60.4	3.6	8095	22	AAK81626
12	60.2	3.6	4858	22	AAQ91904
13	60.2	3.6	4858	24	AAI81674
14	60.2	3.6	1787	21	AAA36622
15	59.8	3.6	5176	22	AAK66814
16	59.8	3.6	56153	22	AAK46793
17	59.6	3.6	19062	21	AAK21280
18	59.6	3.6	23142	21	AAK35158
19	59.6	3.6	30417	21	AAK21282
20	59.6	3.6	30417	21	AAA35160
21	59.2	3.6	5275	22	AAK46378
22	59.2	3.6	5275	22	ABL32825
23	59.2	3.6	160271	22	AAK85750
24	59.2	3.6	160271	22	AAK85756
25	59.2	3.6	160271	22	AAK85756
26	59.2	3.6	160271	22	AAK85756
27	59.2	3.6	160271	22	AAK85756
28	59.2	3.6	160271	22	AAK85756
29	59.2	3.6	160271	22	AAK85756
30	59.2	3.6	160271	22	AAK85756
31	59.2	3.6	160271	22	AAK85756
32	58.8	3.5	3304	21	AAZ45596
33	58.8	3.5	5165	23	ABL26068
34	58.8	3.5	168575	22	AAH21613
35	58.6	3.5	143068	21	AAK21105
36	58.6	3.5	143068	21	AAK21105
37	58.6	3.5	143068	21	AAK21105
38	58.6	3.5	143068	21	AAK21105
39	58.6	3.5	143068	21	AAK21105
40	58.6	3.5	152740	21	AAK21273
41	58.4	3.5	405	22	AAI82412
42	58.4	3.5	437	22	AAI88261
43	58.4	3.5	574	22	ABA59754
44	58.4	3.5	574	22	ABA59754
45	58.4	3.5	574	22	AAK08024

ALIGNMENTS

RESULT 1	AAH43758	AAH43758 standard; DNA; 2170 BP.
XX	AAH43758;	
AC	AAH43758;	
XX		
DT	30-JAN-2002 (first entry)	
XX		
DE	5' flanking and 5' UTR of GLP-2 receptor gene.	
XX		
KW	5' flanking region; 5' UTR; untranslated region; murine;	
KW	glucagon-like peptide-2; GLP-2; receptor; regulator;	
KW	intestinal epithelium; lateral hypothalamus; promoter; ds.	
XX		
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	misc_RNA	1656
FT		/tag= a
FT		/note= "Punitive transcriptional start site"
FT	misc_RNA	1761..1763
FT		/tag= b
FT		/note= "Corresponds to translational start in rat/human
FT		1875..1877
FT	CDS	/tag= c
FT		/note= "Putative translational start site"
PN	W0200179290-A2.	
XX		
XX		
XX		
PD	25-OCT-2001.	
XX		

PF 12-APR-2001; 2001MO-IB00619.
XX
PR 13-APR-2000; 2000US-196909P.
PR 01-FEB-2001; 2001US-265310P.
XX
PA (ONEO-) 1149336 ONTARIO INC.
XX
PI Drucker DJ, Lovshin JAL;
XX
DR WPI; 2002-026012/03.
XX
PT Novel recombinant DNA construct, useful as drug screening tool and as
PT protein delivery vehicle, comprising glucagon-like peptide-2 receptor
PT gene, and linked to heterologous gene of interest for expression -
PS disclosure; Fig 1; 70pp; English.

XX
XX This sequence represents the 5' flanking and 5' UTR of the murine
CC glucagon-like peptide-2 (GLP-2) receptor gene. This sequence may be
CC used in the DNA construct of the invention, such that it is linked
CC for expression with a heterologous gene of interest. This construct
CC is useful for screening compounds to identify regulators of GLP-2
CC receptor expression. The DNA construct is also useful for delivering
CC a gene product to tissues selected from intestinal epithelium and the
CC lateral hypothalamus, by transfecting an organism, or a gametic or
CC embryonic form of such an organism, with the construct. The DNA
CC construct is also useful to identify cells capable of mediating
CC expression from the chosen GLP-2 promoter, to identify regions of
CC the GLP-2 promoter that are functional in a given cell type, and
CC to screen for agents that modulate expression from the GLP-2 promoter.
CC It can be used to map functional regions of the GLP-2 promoter, and
CC for screening agents that modulate the function of GLP-2 promoter.
XX

SO Sequence 2170 BP; 574 A; 501 C; 550 G; 540 T; 5 other:

Query Match 99.9%; Score 1658.8; DB 24; Length 2170;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCTTTACCGGATTTTACACTAGATGCGATCCGCGGAGGTAGCGATGAAGCT 60
DB 1 AGCTTTACCGGATTTTACACTAGATGCGATCCGCGGAGGTAGCGATGAAGCT 60
QY 61 GACCAAGCAAGTCTTTCAGGGGAGAAATCAGAAAAAAATTTTGAAGCAT 120
DB 61 GACCAAGCAAGTCTTTCAGGGGAGAAATCAGAAAAAAATTTTGAAGCAT 120
QY 121 TTCAAGAAAGCAAGATGGAATATTTGTACAAAACAGGTCTTCCGCCCATGCGAC 180
DB 121 TTCAAGAAAGCAAGATGGAATATTTGTACAAAACAGGTCTTCCGCCCATGCGAC 180
QY 181 CGGAGGCTCCGATATGAGCAAGATAGCTTTACAGCTACATTAACAAACACACACAC 240
DB 181 CGGAGGCTCCGATATGAGCAAGATAGCTTTACAGCTACATTAACAAACACACACAC 240
QY 241 ACACACACACACACACACACACACACACACACATGTTTTCTCCCTCCCTCAC 300
DB 241 ACACACACACACACACACACACACACACACACATGTTTTCTCCCTCCCTCAC 300
QY 301 TTCTCTCCATTTCTGTGTGCTCCCAAGAGATGACATATTTAGCTGTAGAAATCACAC 360
DB 301 TTCTCTCCATTTCTGTGTGCTCCCAAGAGATGACATATTTAGCTGTAGAAATCACAC 360
QY 361 CATTAAGGCCATGCGGAGGACATTTCCAGACTGATCTTTTATCATTAAGGTTGAATT 420
DB 361 CATTAAGGCCATGCGGAGGACATTTCCAGACTGATCTTTTATCATTAAGGTTGAATT 420
QY 421 CTTCGCAAGTGTGGATTTTAAGGATTTTAAAGGATTTTATCTAGAGGCACTCAGCTGCT 480
DB 421 CTTCGCAAGTGTGGATTTTAAAGGATTTTAAAGGATTTTATCTAGAGGCACTCAGCTGCT 480
QY 481 CCTGTGATGTTTCAAGATTTACGTGGCTTGTGCTAGCTAATGGAATGATCTATGTTTG 540
DB 481 CCTGTGATGTTTCAAGATTTACGTGGCTTGTGCTAGCTAATGGAATGATCTATGTTTG 540

DB 481 CCTGTGATGTTTCAAGATTTACGTGGCTTGTGCTAGCTAATGGAATGATCTATGTTTG 540
QY 541 ACTTAATGTGAAAGAAAAAGAGGGGAGAAAGAGGAGAGGAGAAAGAGGGAG 600
DB 541 ACTTAATGTGAAAGAAAAAGAGGGGAGAAAGAGGAGAGGAGAAAGAGGGAG 600
QY 601 GGAAAGCTGCTTTTATGCTATTTGCTACTTAACATTTTGTCTCTCAGCTTCCACTTGG 660
DB 601 GGAAAGCTGCTTTTATGCTATTTGCTACTTAACATTTTGTCTCTCAGCTTCCACTTGG 660
QY 661 TTCTTAATGGAAGAGCTGATAGAAAGCTGGAGCCAGCAGGAGATAGAGAGAGTGTGT 720
DB 661 TTCTTAATGGAAGAGCTGATAGAAAGCTGGAGCCAGCAGGAGATAGAGAGAGTGTGT 720
QY 721 GTGTGTGTGGGGGGGGGGTGGGAGCAAGAGAGCTTAGAGACAGAGAGAGAGCTGTAG 780
DB 721 GTGTGTGTGGGGGGGGGGTGGGAGCAAGAGAGCTTAGAGACAGAGAGAGAGCTGTAG 780
QY 781 AGATCATGAGCTTCTTTGAGACCCCTAGTGTACAGAGATAGTCTTAACAGGTAGC 840
DB 781 AGATCATGAGCTTCTTTGAGACCCCTAGTGTACAGAGATAGTCTTAACAGGTAGC 840
QY 841 TGTGTGACGTACTGCTGCTGGAAGSCTGTGCTTTTGTCTGTCTGTGAGCTT 900
DB 841 TGTGTGACGTACTGCTGCTGGAAGSCTGTGCTTTTGTCTGTCTGTGAGCTT 900
QY 901 GAACAAACACCCGCGCTTTTGAACCCCACTATTTCTCAGCCCTCAGATGAAGAATGA 960
DB 901 GAACAAACACCCGCGCTTTTGAACCCCACTATTTCTCAGCCCTCAGATGAAGAATGA 960
QY 961 TGTGTGCTGGAAGATAGTATGAGGTTCAAGTGAAGTGAAGGAGAGGAGGAGGTTTGG 1020
DB 961 TGTGTGCTGGAAGATAGTATGAGGTTCAAGTGAAGTGAAGGAGAGGAGGAGGTTTGG 1020
QY 1021 TAACCAATCACTAAGTGGGGGTGTGTTAGTACTAGTACCAATGCAATTAATATATC 1080
DB 1021 TAACCAATCACTAAGTGGGGGTGTGTTAGTACTAGTACCAATGCAATTAATATATC 1080
QY 1081 TGTGAGTGGCCAGAGCATCTGTGTTCAGAAACACAGGCGCCACACAGAGAGAGGCGGA 1140
DB 1081 TGTGAGTGGCCAGAGCATCTGTGTTCAGAAACACAGGCGCCACACAGAGAGAGGCGGA 1140
QY 1141 CAATTGAAAGGACCAACCTGCTGCTTCTTACCCTGTTTGTGTACCGTGTAAAGCA 1200
DB 1141 CAATTGAAAGGACCAACCTGCTGCTTCTTACCCTGTTTGTGTACCGTGTAAAGCA 1200
QY 1201 ACTCACTCTGCGACATGAAACAGGCTTTGTGTGAGACCTGGGGCTGAGAGTGTGTCT 1260
DB 1201 ACTCACTCTGCGACATGAAACAGGCTTTGTGTGAGACCTGGGGCTGAGAGTGTGTCT 1260
QY 1261 CTGAGACAGAGAAACATCATCTTGTACTATGAGCATAGTAAACACAGAGCTGTGAT 1320
DB 1261 CTGAGACAGAGAAACATCATCTTGTACTATGAGCATAGTAAACACAGAGCTGTGAT 1320
QY 1321 AGCCCTGAGCTGTGCTGCTTTAGAAAAGTTGATGCTTTTAAAGAAATCTGTGCTTAA 1380
DB 1321 AGCCCTGAGCTGTGCTGCTTTAGAAAAGTTGATGCTTTTAAAGAAATCTGTGCTTAA 1380
QY 1381 AGAAGCTTACCTGGCATGGGGGCCATCTCCAGCATCGAATCTCAATCTGTGTGT 1440
DB 1381 AGAAGCTTACCTGGCATGGGGGCCATCTCTCCAGCATCGAATCTCAATCTGTGTGT 1440
QY 1441 GTGCGTAAAGATAGAAATCTCGAATGATTAACATGTCCTTCTTCTGTGGGCTTGC 1500
DB 1441 GTGCGTAAAGATAGAAATCTCGAATGATTAACATGTCCTTCTTCTGTGGGCTTGC 1500
QY 1501 TGAAGAAAGTCCAGGACAGCGTGTGAGCTTTGGGGGTAGGTCTGGGAAAAATCTCCACAGA 1560
DB 1501 TGAAGAAAGTCCAGGACAGCGTGTGAGCTTTGGGGGTAGGTCTGGGAAAAATCTCCACAGA 1560
QY 1561 TTTTAAAGAGGAGGAGGAGGAGGAGTGAAGAACTTGGAGATTCGCTGTAGAGCA 1620
DB 1561 TTTTAAAGAGGAGGAGGAGGAGGAGTGAAGAACTTGGAGATTCGCTGTAGAGCA 1620

OY 1621 ACTCAGACAGTCGGCGCTGAAGAGACTTGTGCAACA 1660
|||||
DB 1621 ACTCAGACAGTCGGCGCTGAAGAGACTTGTGCAACA 1660

RESULT 2

AAQ36051/c
ID AAQ36051 standard; cDNA: 1199 BP.

XX AAQ36051;

XX 24-MAY-1993 (first entry)

XX Human nov gene promoter sequence XXXIII fragment 2.

XX avian nephroblastoma; avian myeloblastoma virus;

XX stringent hybridisation; ss.

XX Homo sapiens.

XX WO9300430-A.

XX 07-JAN-1993.

XX 25-JUN-1992: 92MO-FR00589.

XX 25-JUN-1991: 91FR-0007807.

XX (CNRS) CENT NAT RECH SCI.

XX Martinerie C, Perbal B;

XX WPI: 1993-036377/04.

PT Nucleotide sequences hybridising to regions of chicken nov gene -
PT useful as probes for detecting complementary sequences to
PT evaluate development and/or differentiation of tumours

XX Disclosure; Page 45-46; 67pp; French.

XX The invention includes nucleotide sequences which can hybridise to
XX all or part of the chicken nov gene under stringent conditions
XX (i.e. 50x formamide, 5 x SSC). The promoter sequence from the human
XX nov gene, localised to a 2.2kb PstI-HindII fragment and comprising
XX the 283bp upstream of the start of exon 1 is specifically
XX disclosed. The promoter sequence corresponds to sequence XXXIII
XX which is given as three separate fragments in the specification.
XX The human nov promoter contains several consensus sequences for
XX different transcription factors such as NF1, AP1 and Sp1. It also
XX comprises a sequence of 20 repeats of the TG motif (a possible
XX polymorphic marker). See also AAQ36050 and AAQ36052.

XX Sequence 1199 BP; 269 A; 292 C; 248 G; 390 T; 0 other;

Query Match 3.9%; Score 64.6; DB 14; Length 1199;

Best Local Similarity 57.8%; Pred. No. 2.6e-08; Matches 115; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 79 CAGGGGAGAAAATTCAGAAAAAATTTTGAAGCATTTCAAGACAGATGGA 138
|||||

DB 1025 CAGGGGAGAGGATCTAGAAAAAATCGATATGTGAGAGGTGGGAGATGCTATTGG 966
|||||

OY 139 ATATTTGTCAAAAACAGGTCTTCTCCGCCACATGCGAGCCGGGAGCTCCACTGATAT 198
|||||

DB 965 CCAATATGCAAAATTTCTTACTTAATCCTTGTAGCTTCACTACATCAATTAACCCCA 906
|||||

OY 199 GGAAGATACCTTACGTCATCTCAAAACACACACACACACACACACACACACACAC 258
|||||

DB 905 GGTACACAGTTTGTAGAAATTAAGTAAGACACACACACACACACACACACACACAC 846
|||||

OY 259 AC 277
|||||

DB 845 ACACACACACATTAAGAAA 827

RESULT 3

AAV30918
ID AAV30918 standard; DNA: 245 BP.

XX AAV30918;

XX 14-SEP-1998 (first entry)

XX Human secreted protein AS63_29 cDNA 3' region.

XX AS63_29; secreted protein; protein factor; human; ds.

XX Homo sapiens.

XX WO9817687-A2.

XX 30-APR-1998.

XX 24-OCT-1997: 97MO-US19590.

XX 24-OCT-1997: 97US-0740274.

XX 25-OCT-1996: 96US-0740274.

XX (GENM) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racie LA, Spaulding V, Treacy M;

XX WPI: 1998-261426/23.

PT Nucleic acid encoding secreted protein from human cells - useful,
PT e.g. as immunomodulator, antitumour agent, promoters of tissue
PT growth, haemostatic and thrombolytic agents etc.

XX Claim 16; Page 69; 114pp; English.

XX This polynucleotide comprises the 3' portion of a cDNA clone,
XX designated AS63_29 (see also AAV30917), that codes for a novel human
XX secreted protein (see AAW58381). AS63_29 was isolated from a human
XX foetal brain cDNA library using methods selective for cDNAs that
XX encode secreted proteins. The clone is deposited in composite clone
XX ATCC 98232; an oligonucleotide (see AAT9723) is designed to isolate
XX the clone from the composite. Novel cDNA clones (see AAV30916-32)
XX coding for human secreted proteins (see AAW58580-90) are claimed.
XX These can be used for recombinant production of the secreted
XX proteins for analysis, characterisation, diagnostic or therapeutic
XX use.

XX Sequence 245 BP; 85 A; 69 C; 38 G; 45 T; 8 other;

Query Match 3.8%; Score 63.4; DB 19; Length 245;

Best Local Similarity 77.6%; Pred. No. 2.4e-08; Matches 76; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 220 CATTTTCTTC 289
|||||

DB 73 CATTTTGTATCC 132
|||||

OY 290 CTTCCCTCCACTTCTCTCCATTTCTGTGTCCTCAAG 327
|||||

DB 133 CTTCCTTCCTCCCTCCAGTGGCTTGTGTGATCGCAAG 170
|||||

RESULT 4

AAF98397
ID AAF98397 standard; cDNA: 245 BP.

XX AAF98397;

XX 07-JUN-2001 (first entry)

XX 3' portion of cDNA encoding AS63_29 protein SEQ ID 38.
 DE Human; secreted protein; nutrient; cytokine modulator; proliferation;
 XX differentiation; immune system modulator; tissue growth; chemotactic;
 KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
 KW haematopoiesis.
 XX Homo sapiens.
 OS
 XX MO200119988-A1.
 PN
 XX 22-MAR-2001.
 PD
 XX 14-SEP-2000; 2000MO-US25135.
 PF
 XX 17-SEP-1999; 99US-0398829.
 PR
 XX (GENY) GENETICS INST INC.
 PA
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
 DR WPI; 2001-244801/25.
 XX
 PT Isolated nucleic acids encoding polypeptides, useful for modulating
 PT e.g. cytokine and cell proliferation/differentiation activity, the
 PT immune system and hematopoiesis regulating activity -
 XX
 PS Disclosure: Page 405; 557pp; English.
 XX
 CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
 CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
 CC tissue types, and may be used in the prevention, treatment and diagnosis
 CC of diseases associated with inappropriate protein expression. The
 CC polypeptides and nucleic acids may be used as nutrients or to modulate
 CC cytokine and cell proliferation/differentiation activity and may also be
 CC involved in modulation of the immune system. The cDNA sequences,
 CC proteins, their agonists and/or antagonists exhibit haematopoiesis
 CC regulating activity; tissue growth activity; activin/inhibin activity;
 CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
 CC activity; receptor/ligand activity; anti-inflammatory activity;
 CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
 CC tumour inhibition activity. Included in the invention are probes
 CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
 CC encoding the secreted proteins.
 CC
 XX Sequence 245 BP; 85 A; 69 C; 38 G; 45 T; 8 other;
 SQ
 Query Match 3.8%; Score 63.4; DB 22; Length 245;
 Best Local Similarity 77.6%; Pred. No. 2.4e-08;
 Matches 76; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
 QY 230 CATGTTTCTC 289
 DB 73 CATTTTGATCC 132
 QY 290 CCTCCCTCCACTTCCTCCATTCCTGTGTGTCGCAAG 327
 DB 133 CTTGCTTCCTCCGCGGCGTTCGTGTGATCGCAAG 170
 RESULT 5
 AA188828/c
 ID AA188828 standard; cDNA: 353 BP.
 AC AA188828;
 XX
 XX 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 8888.
 .XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX Homo sapiens.
 OS
 XX MO200164835-A2.
 PN
 XX 07-SEP-2001.
 PD
 XX 26-FEB-2001; 2001MO-US04927.
 PF
 XX 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YL, Liu C, Drmanac RT;
 DR WPI: 2001-514838/56.
 DR P-PSDB: AAO08897.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 1; SEQ ID NO 8888; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 353 BP; 65 A; 59 C; 108 G; 121 T; 0 other;
 SQ
 Query Match 3.7%; Score 61.8; DB 22; Length 353;
 Best Local Similarity 80.9%; Pred. No. 8.7e-08;
 Matches 72; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 226 AAAACACACACACACACACACACACACACACACACACACATGTTT 285
 DB 210 AAAACACACACACACACACACACACACACACACACACACAGGCGCAAGGTTT 151
 QY 286 CTTCCCTCCCTCCACTTCCTCCATTCCTC 314
 DB 150 ATTCCCTCCAGCGCGATTCCTCCACATCTC 122
 RESULT 6
 AAQ12760
 ID AAQ12760 standard; DNA: 3809 BP.
 AC AAQ12760;
 XX
 XX 17-OCT-1991 (first entry)
 DE P40 genomic DNA.
 XX
 XX T cell growth factor; AIDS; cytokine; T helper cell; ss.
 KW
 XX Mus musculus.
 OS
 XX
 FH key Location/Qualifiers

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 6183 BP; 1584 A; 230 C; 1532 G; 2837 T; 0 other;

Query Match 3.7%; Score 60.6; DB 24; Length 6183;

Best Local Similarity 88.0%; Pred. No. 9.6e-07;

Matches 66; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 205 AATAGCTTTACAGTACATTCAAAACACACACACACACACACACACACACACAC 264
DB 3373 AAAACCTACTACTACATTAATAAAAACACACACACACACACACACACACAC 3314
QY 265 ACACACACACACACA 279
DB 3313 ACACACACACACACA 3299

RESULT 8

AAK70103 standard; DNA; 23885 BP.

XX AAK70103;

XX 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24915.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytosolic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX W0200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX
PR 31-JAN-2000; 2000US-01790658.
PR 04-FEB-2000; 2000US-01806528.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 16-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226866.
PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

XX Homo sapiens.
OS WO200164835-A2.
PN 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
XX Tang YF, Liu C, Drmanac RT;
PI WPI: 2001-514838/56.
DR P-PSDB; AA010445.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX Claim 1: SEQ ID NO 10436; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulatory
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 415 BP; 130 A; 109 C; 63 G; 112 T; 1 other;
S0
Query Match 3.6%; Score 60.4; DB 22; Length 415;
Best Local Similarity 74.5%; Pred. No. 2.4e-07;
Matches 76; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 187 CTCACGTATGAGACATAGCTTACAGCTACATTTCAACACACACACACACAC 246
Db 183 CACCTGTCTACTCATATATATAGCTGTAACCTCTTAAGCACACACACACACAC 242
QY 247 ACACACACACACACACACACACACACACATGTTTCTT 288
Db 243 ACACACACACACACACACACACACACACACCTTATT 284
RESULT 11
AAK81626
ID AAK81626 standard; DNA; 8095 BP.
AC AAK81626;
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36438.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.
PF 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205151.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0234000.
PR 14-SEP-2000; 2000US-0234201.
PR 14-SEP-2000; 2000US-0234063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

PA (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lora JM;
XX
XX WPI: 2001-041142/05.
DR P-PSDB: AAB50087.
XX
XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
PT diagnosis of fibrosis, e.g. of the liver
XX
XX
PS Claim 1; Fig 5; 164pp; English.
XX
XX The present sequence is the coding sequence for murine integrin alpha
CC subunit, A259. A259 is homologous with the alpha1 and alpha10 integrin
CC subunits and is overexpressed in fibrosis. A259 is implicated in
CC regulation of proliferation, differentiation and/or function of many
CC different cell types. Inhibitors of A259 activity are useful for the
CC treatment of liver disease, particularly fibrosis, and also fibrosis in
CC other organs (specifically lung and kidney). In addition, A259 can be
CC used for treatment and prevention of cancer, osteoporosis, acute myeloid
CC leukaemia, HIV infection, and rheumatoid arthritis.
XX
XX Sequence 4858 BP; 1184 A; 1397 C; 1296 G; 981 T; 0 other;
SQ
Query Match 3.6%; Score 60.2; DB 22; Length 4858;
Best Local Similarity 68.6%; Pred. No. 1.1e-06;
Matches 83; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
OY 179 CCCGGAGCTCCACTGATATGGACAGAAATAGCTTTACAGCTACATTCAAAACACACACAC 238
DB 4398 CCTGGTAGATGCACTGTGATAGGAGCCCTGTCTTACATCAGACCTACACACACAC 4457
OY 239 ACACACACACACACACACACACACACACACACATGTTTCTCCCTCCCTCC 298
DB 4458 ACATC 4517
OY 299 A 299
DB 4518 A 4518
RESULT 13
AAS16874
ID AAS16874 standard; cDNA; 4858 BP.
XX
AC AAS16874;
XX
DT 14-FEB-2002 (first entry)
XX
DE Murine A259 cDNA.
XX
KW Mouse; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
KW cartilage associated disorder; haematopoietic disorder; bone marrow; ss;
KW immune related disease; apoptotic disorder; neuronal tissue disease;
KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
KW antiarthritic; antinaeemic; antiallergic; antisthmatic; dermatological;
KW antidiabetic; anticonvulsant; antiparkinsonian.
XX
XX Mus musculus.
OS
XX
XX Key location/qualifiers
FH 28..3594
FT CDS
FT /*tag= a
FT /product= "Mouse A259"
FT /note= "This coding sequence is specifically claimed"
FT 28..93
FT sig_peptide
FT /*tag= b
FT 94..3591
FT mat_peptide
FT /*tag= c
FT /product= "Mature murine A259"
XX

PN WO200181414-A2.
XX
XX 01-NOV-2001.
XX
XX 27-APR-2001; 2001MO-US13516.
XX
XX 27-APR-2000; 2000US-0561263.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX Pan Y, Lora J;
XX
XX WPI: 2002-041397/05.
DR P-PSDB: AAU10552.
XX
XX New A259 nucleic acids and polypeptides, which comprise integrin alpha
PT subunit, useful for diagnosing, preventing or treating e.g. liver
PT disease, kidney or lung fibrosis, cancers, blood disorders or immune
PT related diseases
XX
XX Claim 2; Fig 5; 168pp; English.
XX
XX The invention relates to human and murine A259 nucleic acid molecules
CC which encode secreted proteins with homology to integrin alpha subunits,
CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
CC are useful for treating liver disease or fibrosis, particularly kidney
CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
CC useful for diagnosing, preventing or treating cartilage and bone
CC associated disorders (such as bone cancer, achondroplasia, myeloma,
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
CC osteoporosis), bone marrow, blood and haematopoietic disorders (such as
CC acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune
CC related diseases (such as HIV, viral infections, cancers, T cell
CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
CC asthma and psoriasis), apoptotic disorders (such as systemic lupus
CC erythematosus and insulin dependent diabetes mellitus), diseases of the
CC neuronal tissues (such as epilepsy and muscular dystrophy) and
CC neurodegenerative diseases (such as Parkinson's disease and Huntington's
CC disease). This sequence represents cDNA encoding the murine A259
CC polypeptide.
XX
XX Sequence 4858 BP; 1184 A; 1397 C; 1296 G; 981 T; 0 other;
SQ
Query Match 3.6%; Score 60.2; DB 24; Length 4858;
Best Local Similarity 68.6%; Pred. No. 1.1e-06;
Matches 83; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
OY 179 CCCGGAGCTCCACTGATATGGACAGAAATAGCTTTACAGCTACATTCAAAACACACACAC 238
DB 4398 CCTGGTAGATGCACTGTGATAGGAGCCCTGTCTTACATCAGACCTACACACACAC 4457
OY 239 ACACACACACACACACACACACACACACACACATGTTTCTCCCTCCCTCC 298
DB 4458 ACATC 4517
OY 299 A 299
DB 4518 A 4518
RESULT 14
AAA39622
ID AAA39622 standard; DNA; 1787 BP.
XX
AC AAA39622;
XX
DT 01-SEP-2000 (first entry)
XX
DE Murine stomatin DNA.
XX
XX Stomatin; murine; treatment; pain; analgesic; anesthetic; allodynia;
KW diagnostic; ss.
XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:20:41 ; Search time 2257.42 seconds
(without alignments)
15388.369 Million cell updates/sec

Title: US-09-833-740-1_COPY_1_1660

Perfect score: 1660
Sequence: 1 aggttaccgcatttgaca.....gaaagagactgtgcaaca 1660

Scoring table: IDENTITY-NDC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_pl:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	1658.8	99.9	2170	6	AX354807	AX354807 Sequence
2	1658.8	99.9	3130	10	AF338224	AF338224 Mus muscu
3	1601.2	96.5	207160	2	AC016464	AC016464 Mus muscu
4	1601.2	96.5	211173	2	AL646097	AL646097 Mus muscu
5	175	10.5	493	6	AX354812	AX354812 Sequence
6	101.6	6.1	198271	2	AC026591	AC026591 Homo sapi
7	101.6	6.1	210719	2	AC087646	AC087646 Homo sapi
8	101	6.1	213359	2	AC069006	AC069006 Homo sapi
9	77.4	4.7	49665	2	AC100859	AC100859 Homo sapi
10	73.2	4.4	102783	2	AC096120	AC096120 Rattus no
11	72.8	4.4	171322	2	AC095179	AC095179 Rattus no
12	72.2	4.3	72930	2	AC101368	AC101368 Mus muscu
13	71.2	4.3	17529	10	AF417202	AF417202 Mus muscu
14	68.8	4.1	176047	2	AC099092	AC099092 Rattus no
15	68.2	4.1	184550	2	AC095653	AC095653 Rattus no
16	68	4.1	177026	2	AC097947	AC097947 Rattus no
17	67.6	4.1	269035	2	AC093941	AC093941 Rattus no
18	67.4	4.1	74978	2	AC101567	AC101567 Mus muscu
19	67.4	4.1	198696	2	AC087898	AC087898 Mus muscu
20	67.2	4.0	162540	2	AC106235	AC106235 Rattus no
21	67.2	4.0	179604	2	AC099716	AC099716 Mus muscu
22	67	4.0	155258	2	AC093949	AC093949 Rattus no
23	67	4.0	292943	2	AC023605	AC023605 Mus muscu
24	66.8	4.0	447	11	G40821	G40821 Zebraf
25	66.8	4.0	183100	2	AC097616	AC097616 Rattus no
26	66.8	4.0	184884	2	AC105552	AC105552 Rattus no
27	66.6	4.0	284	8	PAB292695	PAB292695 Picea abi
28	66.6	4.0	646	11	G40571	G40571 Zebraf
29	66.6	4.0	16430	10	AB06058852	AB060588 Mus muscu
30	66.6	4.0	121868	2	AC103157	AC103157 Rattus no
31	66.6	4.0	156424	2	AC095092	AC095092 Rattus no
32	66.4	4.0	152316	10	AC007844	AC007844 Mus muscu
33	66.4	4.0	173500	2	AC097832	AC097832 Rattus no
34	66.4	4.0	183665	2	AL645600	AL645600 Mus muscu
35	66.4	4.0	216819	2	AL645725	AL645725 Mus muscu
36	66.2	4.0	140115	2	AC091470	AC091470 Mus muscu
37	66.2	4.0	163177	2	AC106056	AC106056 Rattus no
38	66.2	4.0	198578	2	AC021706	AC021706 Homo sapi
39	66.2	4.0	221511	2	AC087138	AC087138 Mus muscu
40	66.2	4.0	221511	2	AC087556	AC087556 Mus muscu
41	66.2	4.0	222817	2	AL662803	AL662803 Mus muscu
42	66	4.0	364	11	HS210WES	HS210WES H. sapiens
43	66	4.0	107110	2	AC103028	AC103028 Rattus no
44	65.8	4.0	320	11	G45569	G45569 Zebraf
45	65.8	4.0	158949	2	AC094758	AC094758 Rattus no

ALIGNMENTS

RESULT 1
AX354807 LOCUS AX354807 2170 bp DNA linear PAT 06-FEB-2002

DEFINITION Sequence 1 from Patent WO0179290.

ACCESSION AX354807
VERSION AX354807.1 GI:18619538

KEYWORDS

SOURCE synthetic construct.
ORGANISM synthetic construct.
artificial sequence.

REFERENCE

AUTHORS

TITLE

JOURNAL

Patent: WO 0179290-A1 25-Oct-2001;

1149336 ONTARIO INC. (CA)

Location/Qualifiers

1. 2170

FEATURES

source

BASE COUNT 574 a 500 c 531 g 540 t 5 others

ORIGIN

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Recombinant DNA expression construct"

Query Match 99.9%; Score 1658.8; DB 6; Length 2170;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 AGGTTACCGCATTTTACAGCTAGATGGCATCCGTCACGGGTAGCAGGTGATGAAGCT 60
    |||
DB 1 AGGTTACCGCATTTTACAGCTAGATGGCATCCGTCACGGGTAGCAGGTGATGAAGCT 60
OY 61 GACCAAGGCAAGTCTTTCAGGGGGAAGAAATCAGAAAAAATTTTAAAGCAT 120
    |||
DB 61 GACCAAGGCAAGTCTTTCAGGGGGAAGAAATCAGAAAAAATTTTAAAGCAT 120
OY 121 TTCAAGACAGCAAGTGAATATTTGTACAAAACAGGTGCTTCTCCACCATGCGACC 180
    |||
DB 121 TTCAAGACAGCAAGTGAATATTTGTACAAAACAGGTGCTTCTCCACCATGCGACC 180
OY 181 CCGGAGCTCCAGTATGAGCAAGATAGCTTTACAGCTACATTCAAAACACACACAC 240
    |||
DB 181 CCGGAGCTCCAGTATGAGCAAGATAGCTTTACAGCTACATTCAAAACACACACAC 240
OY 241 ACACACACACACACACACACACACACACACACACATGTTTCTCCCTCCAC 300
    |||
DB 241 ACACACACACACACACACACACACACACACACACATGTTTCTCCCTCCAC 300
OY 301 TTCTCCCATTTCTGTGTGCTCCCAAGAGATGACCATTTAGCTGTAGAAATCACACC 360
    |||
DB 301 TTCTCCCATTTCTGTGTGCTCCCAAGAGATGACCATTTAGCTGTAGAAATCACACC 360
OY 361 CATTAAGAGCCATCTGGAGAGCCATTTCCAGACTATCTTTTATCATTTAAAGTTGAAT 420
    |||
DB 361 CATTAAGAGCCATCTGGAGAGCCATTTCCAGACTATCTTTTATCATTTAAAGTTGAAT 420
OY 421 CTTCGACAGTGTGGGTTTAAAGTTTATAGGATTTTATCTAGGGCAGCTACCTGCT 480
    |||
DB 421 CTTCGACAGTGTGGGTTTAAAGTTTATAGGATTTTATCTAGGGCAGCTACCTGCT 480
OY 481 CCCGTGAATTTGAGATTTACTGGGCTGTGTCAGTAAATGAATGATCTATGTTT 540
    |||
DB 481 CCCGTGAATTTGAGATTTACTGGGCTGTGTCAGTAAATGAATGATCTATGTTT 540
OY 541 ACTTAATGTGAAAGAAAAAAGAGGGGAAAAAGAGGAGGAGAGAGAGGAGGAG 600
    |||
DB 541 ACTTAATGTGAAAGAAAAAAGAGGGGAAAAAGAGGAGGAGAGAGAGGAGGAG 600
OY 601 GGAAGAACTGCTTTTATGCTTATGCTACTCTAATCTTTGCTCTACCTTCCATTGG 660
    |||
DB 601 GGAAGAACTGCTTTTATGCTTATGCTACTCTAATCTTTGCTCTACCTTCCATTGG 660
OY 661 TTCTTCAATGGAAGAGCTGATAGAAAGCTGGAGCCAGGAGATAGAGAGAGTGT 720
    |||
DB 661 TTCTTCAATGGAAGAGCTGATAGAAAGCTGGAGCCAGGAGATAGAGAGAGTGT 720
OY 721 GTGTGTGTGGGGGGGGGGGAGCAGACAGAGCCTTAGAGACAGAGAGAGCCTTCTAG 780
    |||
DB 721 GTGTGTGTGGGGGGGGGGGAGCAGACAGAGCCTTAGAGACAGAGAGAGCCTTCTAG 780
OY 781 AGATCATAGCTTCTTTCAGAGCCCTAGTGTCTAACAGAGATAGTCTTAACAGGTAG 840
    |||
DB 781 AGATCATAGCTTCTTTCAGAGCCCTAGTGTCTAACAGAGATAGTCTTAACAGGTAG 840
OY 841 TGTGTACAGTACTGAGCTGGAAGSCTGCTGCTTTTGTCTTGTCTGTGACAGCTT 900
    |||
DB 841 TGTGTACAGTACTGAGCTGGAAGSCTGCTGCTTTTGTCTTGTCTGTGACAGCTT 900
OY 901 GACCAAGACCCCTGCTTTCAGAGCCCTAGTGTCTAACAGAGATAGTCTTAACAGGTAG 960
    |||
DB 901 GACCAAGACCCCTGCTTTCAGAGCCCTAGTGTCTAACAGAGATAGTCTTAACAGGTAG 960
OY 961 TGTGTACAGTACTGAGCTGGAAGSCTGCTGCTTTTGTCTTGTCTGTGACAGCTT 1020
    |||
DB 961 TGTGTACAGTACTGAGCTGGAAGSCTGCTGCTTTTGTCTTGTCTGTGACAGCTT 1020
OY 1021 TAACCATTAAGTGAAGTGGGTGTGTGTTAGTAAAGTAGCCATGAATACATAAATATATC 1080
    |||
DB 1021 TAACCATTAAGTGAAGTGGGTGTGTGTTAGTAAAGTAGCCATGAATACATAAATATATC 1080

```

```

DB 1021 TAACCATTAAGTGAAGTGGGTGTGTGTTAGTAAAGTAGCCATGAATACATAAATATATC 1080
OY 1081 TGTCAAGTGGCCAGAGACATCACTGTCTTACAGAACACAGGCGCCACTCAGAACAGCGGA 1140
    |||
DB 1081 TGTCAAGTGGCCAGAGACATCACTGTCTTACAGAACACAGGCGCCACTCAGAACAGCGGA 1140
OY 1141 CAATTGAAGGACACCAACCTCGTGTCTTCTTACCGCTTGTGTTTGTAAACGTGA 1200
    |||
DB 1141 CAATTGAAGGACACCAACCTCGTGTCTTCTTACCGCTTGTGTTTGTAAACGTGA 1200
OY 1201 ACTCAACTCTCGGCACTGAAACAGCTTTTGTCTCAGACCTGGGGTCTGAGGTGTCT 1260
    |||
DB 1201 ACTCAACTCTCGGCACTGAAACAGCTTTTGTCTCAGACCTGGGGTCTGAGGTGTCT 1260
OY 1261 CTGAGACAGAGAAACATCACTTGTCTTACATAGGCAATAGTACTAACACAGAGCTGTGAT 1320
    |||
DB 1261 CTGAGACAGAGAAACATCACTTGTCTTACATAGGCAATAGTACTAACACAGAGCTGTGAT 1320
OY 1321 AGCCCTGAGCTGTGGCTGTTTAAAGAAATTTGATGCTTTAGAAAGAAATGCTGCTTAAA 1380
    |||
DB 1321 AGCCCTGAGCTGTGGCTGTTTAAAGAAATTTGATGCTTTAGAAAGAAATGCTGCTTAAA 1380
OY 1381 AGAAGCTACCTGAGCTGTGGGGCCCATCTCTCCAGCCATCCGATCTCAATCTGTGCTG 1440
    |||
DB 1381 AGAAGCTACCTGAGCTGTGGGGCCCATCTCTCCAGCCATCCGATCTCAATCTGTGCTG 1440
OY 1441 GTGCGTGAAGATTAAGTATCCCTCGGAATGTGAACATGTCCTTTCTTCTGTGGGCTTGC 1500
    |||
DB 1441 GTGCGTGAAGATTAAGTATCCCTCGGAATGTGAACATGTCCTTTCTTCTGTGGGCTTGC 1500
OY 1501 TGAGAGAGTCCAGAGCAGCTAGACGTCTTGGGGGTAGTCTGGGAAAAATCTCCCAAGA 1560
    |||
DB 1501 TGAGAGAGTCCAGAGCAGCTAGACGTCTTGGGGGTAGTCTGGGAAAAATCTCCCAAGA 1560
OY 1561 TTTTGAAGGGGAGAGCGGGGGGATGAGAACTTGGAGATTCGTGATAGCA 1620
    |||
DB 1561 TTTTGAAGGGGAGAGCGGGGGGATGAGAACTTGGAGATTCGTGATAGCA 1620
OY 1621 ACTCAGACAGTGGGCGCTGAAAGAGACTTGTCAAA 1660
    |||
DB 1621 ACTCAGACAGTGGGCGCTGAAAGAGACTTGTGTCAAA 1660

```

RESULT 2
AF338224 3130 bp DNA linear ROD 18-JUN-2001
LOCUS
DEFINITION
AF338224 Mus musculus glucagon-like peptide-2 receptor gene, partial cds.
VERSION
AF338224.1 GI:14485646
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3130)
Lovshin,J.A., Estall,J., Yusta,B., Brown,T.J. and Drucker,D.J.
Glucagon-like peptide (GLP)-2 Action in the Murine Central Nervous
System Is Enhanced by Elimination of GLP-1 Receptor Signaling
J. Biol. Chem. 276 (24), 21489-21499 (2001)
JOURNAL
PUBMED
11262390
REFERENCE
2 (bases 1 to 3130)
Lovshin,J.A. and Drucker,D.J.
Direct Submission
Submitted (17-JAN-2001) Department of Medicine, University of
Toronto, 101 College Street CCRW3-845, Toronto, ON M5G2C4, Canada
location/Qualifiers
1..3130
/organism="Mus musculus"
/strain="129/SvJ"
/db_xref="taxon:10090"
2617..2900
/number=1
exon

```

mRNA      2617. .>2900
/product="glucagon-like peptide-2 receptor"
CDS       2835. >2900
          /note="GLP-2 receptor"
          /codon_start=1
          /product="glucagon-like peptide-2 receptor"
          /protein_id="AAK63043.1"
          /db_xref="GI:14485647"
          /translation="MRRLMGPEPPLSLLLVSIK"
BASE COUNT      805 a      741 c      769 g      797 t      18 others
ORIGIN
Query Match      99.9%; Score 1658.8; DB 10; Length 3130;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGGTTATCCGCATTTTACACTAGATGCGATCCGTCACGGGTAGCAGTCAAGCT 60
          |||||||
DB      961 AGGTTATCCGCATTTTACACTAGATGCGATCCGTCACGGGTAGCAGTCAAGCT 1020

OY      61 GACCAAGGCAAGCTCTTTCAGGGGGAAGAAATCAGAAAAAATTTTGAAGCAT 120
          |||||||
DB      1021 GACCAAGGCAAGCTCTTTCAGGGGGAAGAAATCAGAAAAAATTTTGAAGCAT 1080

OY      121 TTCAAGAAACCAAGATGGAATATTTTACAAAACAGGTCTTCTCCCCACATGCGACC 180
          |||||||
DB      1081 TTCAAGAAACCAAGATGGAATATTTTACAAAACAGGTCTTCTCCCCACATGCGACC 1140

OY      181 CGGAGCTCCACTGATATGAGACAGAAATAGCTTTACAGCTACATTCAAAACACACACAC 240
          |||||||
DB      1141 CGGAGCTCCACTGATATGAGACAGAAATAGCTTTACAGCTACATTCAAAACACACACAC 1200

OY      241 ACACACACACACACACACACACACACACACACACACATTTTCTCCCTCCCTCCAC 300
          |||||||
DB      1201 ACACACACACACACACACACACACACACACACACACATTTTCTCCCTCCCTCCAC 1260

OY      301 TTCTCCCATCTCTGTGTGCTCCCAAGAGATGACCATATTTGACTGTGAATACACACAC 360
          |||||||
DB      1261 TTCTCCCATCTCTGTGTGCTCCCAAGAGATGACCATATTTGACTGTGAATACACACAC 1320

OY      361 CATAAAAAGCCCATCTGGAGCCATTTCCAGACATCTTTTATCATTTAAAGTTTGAAAT 420
          |||||||
DB      1321 CATAAAAAGCCCATCTGGAGCCATTTCCAGACATCTTTTATCATTTAAAGTTTGAAAT 1380

OY      421 CTTCGCCACGTGTGGGTTTAAAGTTTAAAGTATTTTATCTACGCGACATCACTGCTT 480
          |||||||
DB      1381 CTTCGCCACGTGTGGGTTTAAAGTATTTTAAAGTATTTTATCTACGCGACATCACTGCTT 1440

OY      481 CCTGTGAATGTTCAGAAATCACTGGGCTGTGCAGCTATATGAAATGATCTATGTTTG 540
          |||||||
DB      1441 CCTGTGAATGTTCAGAAATCACTGGGCTGTGCAGCTATATGAAATGATCTATGTTTG 1500

OY      541 ACTTAATGTGAAGAAAAAAGAGGGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
          |||||||
DB      1501 ACTTAATGTGAAGAAAAAAGAGGGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560

OY      601 GGAAGACGCTCTTATGCTATGCTACTCTAATCATTTTGTCTCTACCTTCCACTTGG 660
          |||||||
DB      1561 GGAAGACGCTCTTATGCTATGCTACTCTAATCATTTTGTCTCTACCTTCCACTTGG 1620

OY      661 TTCTTAATGAAGAACTGTATGAAAGCTGGAGCCAGCCAGGATGAGAGAGTGTGT 720
          |||||||
DB      1621 TTCTTAATGAAGAACTGTATGAAAGCTGGAGCCAGCCAGGATGAGAGAGTGTGT 1680

OY      721 GTGTGTGTGGGGGGGGTGGGAGCAAGACAGACCTTTAGACAGACAGAGACCTGCTAG 780
          |||||||
DB      1681 GTGTGTGTGGGGGGGGTGGGAGCAAGACAGACCTTTAGACAGACAGAGACCTGCTAG 1740

OY      781 AGAYCAATGAGCTTCTTTGAGACCCCTAGTGTCAACAGAAATGTTCTTCAACAGGTAGC 840
          |||||||
DB      1741 AGAYCAATGAGCTTCTTTGAGACCCCTAGTGTCAACAGAAATGTTCTTCAACAGGTAGC 1800

OY      841 TGTGTACAGTGAAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCTT 900

```

```

|||||
DB      1801 TGTGTACAGTGAAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCTT 1860

OY      901 GAACAAACACCTTGGGCTCTTTAACCCTCAATTTTCTGAGCCCTCAGATGAAGAAATA 960
          |||||||
DB      1861 GAACAAACACCTTGGGCTCTTTAACCCTCAATTTTCTGAGCCCTCAGATGAAGAAATA 1920

OY      961 TGTACCTTGGAGATACATGATGATGTTTCAAGTAACTAGAGGAGAGGAGTGAAGTGTG 1020
          |||||||
DB      1921 TGTACCTTGGAGATACATGATGATGTTTCAAGTAACTAGAGGAGAGGAGTGAAGTGTG 1980

OY      1021 TAAACATTAACCTGAAGTGGGCTTGTGTTAGTAACTAGACCATGAATACATAAATATC 1080
          |||||||
DB      1981 TAAACATTAACCTGAAGTGGGCTTGTGTTAGTAACTAGACCATGAATACATAAATATC 2040

OY      1081 TGTACAGTGGCCAGAGCATATCTGTGTTCAAGAACACACACAGGCCACTCACAACCGGGA 1140
          |||||||
DB      2041 TGTACAGTGGCCAGAGCATATCTGTGTTCAAGAACACACACAGGCCACTCACAACCGGGA 2100

OY      1141 CAATGAAAGGACCAACCTCCGCTCTCCACCCGTTGTTTGTACCGTGTAAACGCA 1200
          |||||||
DB      2101 CAATGAAAGGACCAACCTCCGCTCTCCACCCGTTGTTTGTACCGTGTAAACGCA 2160

OY      1201 ACTCACTCTCGGCACCTGAACAGGCTTTTCTGCAAGCTGGGGTCTGAGGTGTGCT 1260
          |||||||
DB      2161 ACTCACTCTCGGCACCTGAACAGGCTTTTCTGCAAGCTGGGGTCTGAGGTGTGCT 2220

OY      1261 CTGAGACAGAAAACTCATCTTTTACTATAGCATATGATTAACACAGGAGCTGTGAT 1320
          |||||||
DB      2221 CTGAGACAGAAAACTCATCTTTTACTATAGCATATGATTAACACAGGAGCTGTGAT 2280

OY      1321 AGCCCTGAGCTGGGCGCTTTAGAAAAGTTGATGCTTTGAAAAGAAATGTGCTTAAA 1380
          |||||||
DB      2281 AGCCCTGAGCTGGGCGCTTTAGAAAAGTTGATGCTTTGAAAAGAAATGTGCTTAAA 2340

OY      1381 AGAAGCTACCTGGCATGGGGGCCATCTCTCCAGCCATCCGAATCAATCTGTGCT 1440
          |||||||
DB      2341 AGAAGCTACCTGGCATGGGGGCCATCTCTCCAGCCATCCGAATCAATCTGTGCT 2400

OY      1441 GTGCGTAAAGATAGATATCTCTGGAATGTAAACATGCTTCTTCTCTGCGGCTTGC 1500
          |||||||
DB      2401 GTGCGTAAAGATAGATATCTCTGGAATGTAAACATGCTTCTTCTCTGCGGCTTGC 2460

OY      1501 TGAAGAGTCCCGACGCGCTGAGACGCTTGGGGGTAGGTCTGGAAAAAATCTCCCAAGA 1560
          |||||||
DB      2461 TGAAGAGTCCCGACGCGCTGAGACGCTTGGGGGTAGGTCTGGAAAAAATCTCCCAAGA 2520

OY      1561 TTTTAGAGGGGCGAGCGGGGGATGAGAAACTTGGAGATTCGATAGCTGCTAGAGCA 1620
          |||||||
DB      2521 TTTTAGAGGGGCGAGCGGGGGATGAGAAACTTGGAGATTCGATAGCTGCTAGAGCA 2580

OY      1621 ACTCAGACATGCGCGGCTGTAAGAGACCTTGTCAAAACA 1660
          |||||||
DB      2581 ACTCAGACATGCGCGGCTGTAAGAGACCTTGTCAAAACA 2620

RESULT 3
AC016464      207160 bp      DNA      linear      HTG 01-MAR-2000
LOCUS      AC016464
DEFINITION      Mus musculus chromosome 11 clone RP23-40921 map 11, WORKING DRAFT
ACCESSION      AC016464
VERSION      AC016464.3 GI:7137319
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM      Mus musculus.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 207160)
JOURNAL      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE      Mus musculus chromosome 11, clone RP23-40921
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 207160)

```

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckhalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gaityna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Liu, C., Locke, K., Macdonald, P., Marcuis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meidrin, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA.
On Mar 1, 2000 this sequence version replaced g1:6970326.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L3198

Center clone name: 409_J_21

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 196848 bases at least Q40
Consensus quality: 202204 bases at least Q30
Consensus quality: 204082 bases at least Q20
Insert size: 21000; agarose-IP
Insert size: 205160; sum-of-coverage
Quality coverage: 4.8 in Q20 bases; agarose-IP
Quality coverage: 4.9 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 541: contig of 541 bp in length
* 542 641: gap of 100 bp
* 642 1859: contig of 1218 bp in length
* 1860 1959: gap of 100 bp
* 1960 4152: contig of 2193 bp in length
* 4153 4252: gap of 100 bp
* 4253 6291: contig of 2039 bp in length
* 6292 6391: gap of 100 bp
* 6392 10537: contig of 4146 bp in length
* 10538 10637: gap of 100 bp
* 10638 14144: contig of 3507 bp in length
* 14145 14244: gap of 100 bp
* 14245 18479: contig of 4235 bp in length
* 18480 18579: gap of 100 bp
* 18580 25538: contig of 6959 bp in length
* 25539 25638: gap of 100 bp
* 25639 31995: contig of 6357 bp in length
* 31996 32095: gap of 100 bp
* 32096 40133: contig of 8038 bp in length
* 40134 40233: gap of 100 bp
* 40234 48264: contig of 8031 bp in length
* 48265 48364: gap of 100 bp
* 48365 55351: contig of 6987 bp in length
* 55352 55451: gap of 100 bp

FEATURES

source

* 55452 63846: contig of 8395 bp in length
* 63847 63946: gap of 100 bp
* 63947 75456: contig of 11510 bp in length
* 75457 75556: gap of 100 bp
* 75557 88148: contig of 12592 bp in length
* 88149 88248: gap of 100 bp
* 88249 99481: contig of 11233 bp in length
* 99482 99581: gap of 100 bp
* 99582 115973: contig of 16392 bp in length
* 115974 116073: gap of 100 bp
* 116074 132294: contig of 16221 bp in length
* 132295 132394: gap of 100 bp
* 132395 149262: contig of 16868 bp in length
* 149263 149362: gap of 100 bp
* 149363 177577: contig of 28215 bp in length
* 177578 177677: gap of 100 bp
* 177678 207160: contig of 29483 bp in length.
Location/Qualifiers
1. 207160
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-409J21"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 541
/note="assembly-fragment
clone_end:77
vector_side:right"
642. 1859
/note="assembly-fragment"
1960. 4152
/note="assembly-fragment"
4253. 6291
/note="assembly-fragment"
6392. 10537
/note="assembly-fragment"
10638. 14144
/note="assembly-fragment"
14245. 18479
/note="assembly-fragment"
18580. 25538
/note="assembly-fragment"
25639. 31995
/note="assembly-fragment"
32096. 40133
/note="assembly-fragment"
40234. 48264
/note="assembly-fragment"
48365. 55351
/note="assembly-fragment"
55452. 63846
/note="assembly-fragment"
63947. 75456
/note="assembly-fragment"
75557. 88148
/note="assembly-fragment"
88249. 99481
/note="assembly-fragment"
99582. 115973
/note="assembly-fragment"
116074. 132294
/note="assembly-fragment"
132395. 149262
/note="assembly-fragment"
149363. 177577
/note="assembly-fragment"
177578. 177677
/note="assembly-fragment"
177678. 207160
/note="assembly-fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 57689 a 46293 c 45874 g 55300 t 2004 others
ORIGIN

Query Match 96.5%; Score 1601.2; DB 2; Length 207160;
 Best local similarity 99.0%; Pred. No. 0;
 Matches 161; Conservative 3; Mismatches 6; Indels 8; Gaps 4;

QY 1 AGTTTACCCGATTTTGACACTAGATGGCATCCGTCCAGGGGTAGAGGTCATGAAGCT 60
 |||||||
 Db 27190 AGGTTTACCCGATTTTGACACTAGATGGCATCCGTCCAGGGGTAGAGGTCATGAAGCT 27249

QY 61 GACCAAGGCAAGTCCTTTCAGGGGGAAGAAATCAGG--AAAAAATTTTGAAGC 118
 |||||||
 Db 27250 GACCAAGGCAAGTCCTTTCAGGGGGAAGAAATCAGGAAAAAATTTTGAAGC 27309

QY 119 ATTTCAGAAAGCAAGATGAAATATTTGTACAAAACAGGTCTTCTCCACCATCGCA 178
 |||||||
 Db 27310 ATTTCAGAAAGCAAGATGAAATATTTGTACAAAACAGGTCTTCTCCACCATCGCA 27369

QY 179 CCGGGAGCT--CCACGATATGAGCAAGATAGCTTACAGCTCATTCANA--ACACAC 234
 |||||||
 Db 27370 CCGGGAGCTCCACGATATGAGCAAGATAGCTTACAGCTCATTCANAACACACAC 27429

QY 235 ACATGTTTCTCCCTCC 294
 |||||||
 Db 27430 ACACACACACACACACACACACACACACACACACACACATGTTTCTCCCTCC 27489

QY 295 CTCACCTTCTCCCTATCTCTGTGTGCTCCAAAGATGACATATGCTGTAGAAATCA 354
 |||||||
 Db 27490 CTCACCTTCTCCCTATCTCTGTGTGCTCCAAAGATGACATATGCTGTAGAAATCA 27549

QY 355 CACACCATTAAGCCCATCTGGGAGCCATTCAGACTGATCTTTATCATTAAGGT 414
 |||||||
 Db 27550 CACACCATTAAGCCCATCTGGGAGCCATTCAGACTGATCTTTATCATTAAGGT 27609

QY 415 TGAATCTTGCACGTGTGGGTTTAAAGTTTAAAGATTTTATCTAGCGGCACTAC 474
 |||||||
 Db 27610 TGAATCTTGCACGTGTGGGTTTAAAGTTTAAAGATTTTATCTAGCGGCACTAC 27669

QY 475 CTGCTTCCCTGATATGTCGAATTCACCTGGCTTGGACGCTAATGAAATGATCTAT 534
 |||||||
 Db 27670 CTGCTTCCCTGATATGTCGAATTCACCTGGCTTGGACGCTAATGAAATGATCTAT 27729

QY 535 GGTTCGACTTAATGTGAAGGAAAAAAGAAAGGAGGAAAAAGAGGAGGAGAAAG 594
 |||||||
 Db 27730 GGTTCGACTTAATGTGAAGGAAAAAAGAAAGGAGGAAAAAGAGGAGGAGAAAG 27789

QY 595 GGGAGGAAAACTGCCCTTTATGCCATTTGCTACTTAACATTTTGTCTCACCCTTC 654
 |||||||
 Db 27790 GGGAGGAAAACTGCCCTTTATGCCATTTGCTACTTAACATTTTGTCTCACCCTTC 27849

QY 655 ACTTGCTCTTCAATGGAAGACTGGATGAAGGCGGAGGAGCAGGATAGAGGA 714
 |||||||
 Db 27850 ACTTGCTCTTCAATGGAAGACTGGATGAAGGCGGAGGAGCAGGATAGAGGA 27909

QY 715 --GTGTGTGTGTGTGGGGGGGGGTGGGACAGCAAGCAGAGCCTTAGAGACAGAGAG 772
 |||||||
 Db 27910 GTGTGTGTGTGTGTGGGGGGGGGTGGGACAGCAAGCAGAGCCTTAGAGACAGAGAG 27969

QY 773 CCTGCTAGATCATGAGCTTCTTTGAARACCCCTAGTCTAAGAGATAGTCTTAC 832
 |||||||
 Db 27970 CCTGCTAGATCATGAGCTTCTTTGAARACCCCTAGTCTAAGAGATAGTCTTAC 28029

QY 833 CAGGTAGCTGTGCTGACGTGACTCGGCTGGAAGSCTGGCTTGTCTTTTGGCTCTGT 892
 |||||||
 Db 28030 CAGGTAGCTGTGCTGACGTGACTCGGCTGGAAGSCTGGCTTGTCTTTTGGCTCTGT 28089

QY 893 GCAGCCTTGAACAAACCCCTGGCTCTTTGAACCCCACTATTCTTCAGCCCTCAATGA 952
 |||||||
 Db 28090 GCAGCCTTGAACAAACCCCTGGCTCTTTGAACCCCACTATTCTTCAGCCCTCAATGA 28149

QY 953 AGAAGTAATGTACTCTTGAAGATGATGATGGTTCAAGTGAAGTGAAGGAGAGGAGTGA 1012
 |||||||
 Db 28150 AGAAGTAATGTACTCTTGAAGATGATGATGGTTCAAGTGAAGTGAAGGAGAGGAGTGA 28209

QY 1013 AGTTTTGTAAACATAAATGAAGTGGGTGTGGTGTAGTAAGTAGCCATGAATACCATA 1072
 |||||||
 Db 28210 AGTTTTGTAAACATAAATGAAGTGGGTGTGGTGTAGTAAGTAGCCATGAATACCATA 28269

QY 1073 AAAATATCTGTCAAGTGGCCAGACATCACTGTGTTAGAAACAAAGGCCCACTCGAA 1132
 |||||||
 Db 28270 AAAATATCTGTCAAGTGGCCAGACATCACTGTGTTAGAAACAAAGGCCCACTCGAA 28329

QY 1133 CACGGGCAATTAAGAGGACCAACCTCCGTCTTACCCGTTGTTTGAACGGTG 1192
 |||||||
 Db 28330 CACGGGCAATTAAGAGGACCAACCTCCGTCTTACCCGTTGTTTGAACGGTG 28389

QY 1193 TAAACGCACTCACTCTCGGCACTGAACAGCTTTTGTCTGCAAGCTGGGCTCGAAG 1252
 |||||||
 Db 28390 TAAACGCACTCACTCTCGGCACTGAACAGCTTTTGTCTGCAAGCTGGGCTCGAAG 28449

QY 1253 TGTGTCTCTGAGACAGAAACATCTTGTCTTACTATGCAATGATAGTAAACAGGAGC 1312
 |||||||
 Db 28450 TGTGTCTCTGAGACAGAAACATCTTGTCTTACTATGCAATGATAGTAAACAGGAGC 28509

QY 1313 TCTGATAGCCCTGAGCTGTGCTTGAAGAAAGTTGATGCTTTGAAAGAAATCGT 1372
 |||||||
 Db 28510 TCTGATAGCCCTGAGCTGTGCTTGAAGAAAGTTGATGCTTTGAAAGAAATCGT 28569

QY 1373 GGCCTTAAAGAGCCCTACCTGGCATGGGGCCATCTCTCCAGCATCCGAATCTCAAT 1432
 |||||||
 Db 28570 GGCCTTAAAGAGCCCTACCTGGCATGGGGCCATCTCTCCAGCATCCGAATCTCAAT 28629

QY 1433 CTGCTGTGTGCTGAAGATGAATCTCGGAATGATGATGATGATGATGATGATGAT 1492
 |||||||
 Db 28630 CTGCTGTGTGCTGAAGATGAATCTCGGAATGATGATGATGATGATGATGATGAT 28689

QY 1493 GGGCTTGCAGAGAGTCCACAGGAGCGTAGACGCTTGGGGGTAGTCTGGGAAAAATC 1552
 |||||||
 Db 28690 GGGCTTGCAGAGAGTCCACAGGAGCGTAGACGCTTGGGGGTAGTCTGGGAAAAATC 28749

QY 1553 TCCCAAGATTTTGAAGAGGCGAGCGGGGGATGAGAAACTTGGAGATTCGGTATCGCT 1612
 |||||||
 Db 28750 TCCCAAGATTTTGAAGAGGCGAGCGGGGGATGAGAAACTTGGAGATTCGGTATCGCT 28809

QY 1613 GTAGAGCAACTGACAGACAGTCGGCGCTGAAGAGACTTGTCCAAACA 1660
 |||||||
 Db 28810 GTAGAGCAACTGACAGACAGTCGGCGCTGAAGAGACTTGTGTCAACA 28857

RESULT 4
 AL646097/3 211173 bp DNA linear HTG 10-JAN-2002
 LOCUS Mus musculus chromosome 11 clone RP23-338M9, *** SEQUENCING IN
 DEFINITION
 ACCESSION AL646097 GI:18135259
 VERSION AL646097.5
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 McIlroy, K.
 DIRECT SUBMISSION
 Submitted (08-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Jan 11, 2002 this sequence version replaced gi:18072579.
 Genomes Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 Project Information
 Center project name: BM318M9
 Summary Statistics
 Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 206626 bases at least Q40
Consensus quality: 207698 bases at least Q30
Consensus quality: 208567 bases at least Q20
Insert size: 209573; sum-of-contigs
Insert size: 207120; 2.8% error; agarose-fp
Quality coverage: 9.05x in Q20 bases; sum-of-contigs quality
coverage: 9.26x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

source

Location/Qualifiers

1. 211173

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="11"

/clone="RP23-338M9"

/clone_11b="RPC1-23"

1. 8490

misc_feature
/note="assembly_fragment:04365
fragment_chain:1"

misc_feature
/note="assembly_fragment:02882
fragment_chain:1"

misc_feature
/note="assembly_fragment:01376
fragment_chain:1"

misc_feature
/note="assembly_fragment:01658
fragment_chain:1"

misc_feature
/note="assembly_fragment:00263
fragment_chain:1"

misc_feature
/note="assembly_fragment:02994
fragment_chain:2"

misc_feature
/note="assembly_fragment:00672
fragment_chain:2"

misc_feature
/note="assembly_fragment:03096
fragment_chain:2"

misc_feature
/note="assembly_fragment:01861
fragment_chain:2"

misc_feature
/note="assembly_fragment:01143
fragment_chain:2"

misc_feature
/note="assembly_fragment:00621
fragment_chain:2"

misc_feature
/note="assembly_fragment:00072
fragment_chain:2"

misc_feature
/note="assembly_fragment:01439
fragment_chain:2"

misc_feature
/note="assembly_fragment:00832
fragment_chain:2"

misc_feature
/note="assembly_fragment:03362
fragment_chain:2"

misc_feature
/note="assembly_fragment:03683
fragment_chain:2"

misc_feature
/note="assembly_fragment:01683
fragment_chain:2"

misc_feature
/note="assembly_fragment:01683
fragment_chain:2"

misc_feature
/note="assembly_fragment:01683
fragment_chain:2"

BASE COUNT 57360 a 48310 c 47762 g 56133 t 1608 others
ORIGIN

Query Match 96.5%; Score 1601.2; DB 2; Length 211173;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1651; Conservative 3; Mismatches 6; Indels 8; Gaps 4;

QY 1 AGTTTACCGCATTTTGCACATAGATGCGATCCGCCAGGGGTGACGATGATGACT 60
Db 133842 AGGTTTACCGCATTTTGCACATAGATGCGATCCGCCAGGGGTGACGATGATGACT 133783
QY 61 GACCAAGGCAAGTCTTTACAGGGGGAGAGAAATCAG--AAAAAATTTTGAAGC 118
Db 133782 GACCAAGGCAAGTCTTTACAGGGGGAGAGAAATCAGAAAAAATTTTGAAGC 133723
QY 119 ATTTCAGAGCAAGATGGAATTTTGTACAAACAGTGTCTTCCGCCACATGCGA 178
Db 133722 ATTTCAGAGCAAGATGGAATTTTGTACAAACAGTGTCTTCCGCCACATGCGA 133663
QY 179 CCGGGAGCT--CCAGTATATGACAGAAATAGCTTTACAGCTACATTCANA--ACACAC 234
Db 133662 CCGGGAGCTCCCGACATGATATGATATTTGACAAACAGTGTCTTCCGCCACATGCGA 133603
QY 235 ACATGTTTCTCTCC 294
Db 133602 ACATGTTTCTCTCC 133543
QY 295 CTCACCTTCTCCATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 354
Db 133542 CTCACCTTCTCCATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 133483
QY 355 CACACACATTAAGCCATCTGGAGCCATTTCCAGCATCTTTTATCATTAAGTT 414
Db 133482 CACACACATTAAGCCATCTGGAGCCATTTCCAGCATCTTTTATCATTAAGTT 133423
QY 415 TGAATCTTCCACGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 474
Db 133422 TGAATCTTCCACGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 133363
QY 475 CTGCTTCCCTGTGAATGTTGATTCATCTGAGCTGTTGATTCATCTGATTCAT 534
Db 133362 CTGCTTCCCTGTGAATGTTGATTCATCTGAGCTGTTGATTCATCTGATTCAT 133303
QY 535 GGTGTACCTTAATGTGAAGGAAAAAGAGGGGAAAAAGAGGAGGAGGAGAGAG 594
Db 133302 GGTGTACCTTAATGTGAAGGAAAAAGAGGGGAAAAAGAGGAGGAGGAGAGAG 133243
QY 595 GGGAAAGGAAAACTGCTTTTATGCTATTTGCTACTTAATTTTGTCTGACCTGC 654
Db 133242 GGGAAAGGAAAACTGCTTTTATGCTACTTAATTTTGTCTGACCTGC 133183
QY 655 ACTTGTCTTCTTAATGGAAGAGCTGATGAAAGCTGGAGCCAGCCAGGATAGAGGA 714
Db 133182 ACTTGTCTTCTTAATGGAAGAGCTGATGAAAGCTGGAGCCAGCCAGGATAGAGGA 133123
QY 715 --GTGTGTGTGTGTGGGGGGGGGGGGGAGAGCAAGCAAGCCTTAGAGAGAGAG 772
Db 133122 GTGTGTGTGTGTGTGGGGGGGGGGGGGAGAGCAAGCAAGCCTTAGAGAGAGAG 133063
QY 773 CCTGCTAGAGATCATGAGCTTCTTGAAGCCCTAGTCTTAACAGGAATAGTCTTAAC 832
Db 133062 CCTGCTAGAGATCATGAGCTTCTTGAAGCCCTAGTCTTAACAGGAATAGTCTTAAC 133003
QY 833 CAGGTAGCTGT 892
Db 133002 CAGGTAGCTGT 132943
QY 893 GCAGCCTTGAACAAACACCTGGCTTTTGAACCCCATTTTTCAGCCCTTGAGATGA 952
Db 132942 GCAGCCTTGAACAAACACCTGGCTTTTGAACCCCATTTTTCAGCCCTTGAGATGA 132883
QY 953 AGAAGTAATGCTTCTGAGAGATATCTATGCTTCAAGTGAATAGGAGAGAGGTGA 1012

	D	b	132882	AGAAATTAATGGTTACTTGTCGACGATACTACGTATCGGTTTCAAGTAACATTAGGCACAGGGTGGGA	132823
Oy	1013	AGTTTTTGTAAACAATPAAACCTGAAGTGGGGTTGTGTTAAGTAAGTAGCCATGAATACCATA	1072		
D	132822	AGGTTTGTGAACCAATPAAACTGAATGGGGTTGTGTTAAGTAAGTAGCCATGAATACCATA	132763		
Oy	1073	AAAATATCTGCAGGTGGGCCAGACATCAGCTGTCTGAGAACAACAGGCCCATCTCGAAA	1132		
D	132762	AAAATATCTGCAGGTGGGCCAGACATCAGCTGTCTGAGAACAACAGGCCCATCTCGAAA	132703		
Oy	1133	CACGGGACAATTTGAAGGACCAACAACCTCGTGCTTCCTACCCGTGTTTGTAAACCGTGC	1192		
D	132702	CACGGGACAATTTGAAGGACCAACAACCTCGTGCTTCCTACCCGTGTTTGTAAACCGTGC	132643		
Oy	1193	TAAACGCAACTCACTCTCGGCACATGAACAGGCTTTGCTCAGACCCGAGGGGTCTGGAGG	1252		
D	132642	TAAACGCAACTCACTCTCGGCACATGAACAGGCTTTGCTCAGACCCGAGGGGTCTGGAGG	132583		
Oy	1253	TGTTGCTCTGTAGACAGGAAAAACTCATCTTTGTACTATGAGCATAAGTAGTAACAAGGAGC	1312		
D	132582	TGTTGCTCTGTAGACAGGAAAAACTCATCTTTGTACTATGAGCATAAGTAGTAACAAGGAGC	132523		
Oy	1313	TCTGAGATAGCCCTCAGCTGCTGGCCGTTTAAGAAAAGTTGATTCGTTTGAAGAAAATTCG	1372		
D	132522	TCTGAGATAGCCCTCAGCTGCTGGCCGTTTAAGAAAAGTTGATTCGTTTGAAGAAAATTCG	132463		
Oy	1373	GGCCTTAAAAAGAACCTTACCTGGCATGGGGGCCATCCTCTCACGCAATCCGAATCTCAAT	1432		
D	132462	GGCCTTAAAAAGAACCTTACCTGGCATGGGGGCCATCCTCTCACGCAATCCGAATCTCAAT	132403		
Oy	1433	CTGTGCTGTGTGCGTAAAGATAGATATCTCTGGAAATGTAACCATGTGCTTTTCTTCTCT	1492		
D	132402	CTGTGCTGTGTGCGTAAAGATAGATATCTCTGGAAATGTAACCATGTGCTTTTCTTCTCT	132343		
Oy	1493	GGGCTTGTGGAAGAAAGTCCCAGAGCGATGACGCTCTTGSGGGGTAGGCTGGGAAAAATTC	1552		
D	132342	GGGCTTGTGGAAGAAAGTCCCAGAGCGATGACGCTCTTGSGGGGTAGGCTGGGAAAAATTC	132283		
Oy	1553	TCCCAAGATTTTtagGAGGggcAgcgGGgGGATGaGAAAaCTtGGAGAtttOGGTAgTCGCT	1612		
D	132282	TCCCAAGATTTTtagGAGGggcAgcgGGgGGATGaGAAAaCTtGGAGAtttOGGTAgTCGCT	132223		
Oy	1613	GTAGAGCAACTCAGACAGTCCGGGGCGCTCAAGAGCACTGTGCAACA	1660		
D	132222	GTAGAGCAACTCAGACAGTCCGGGGCGCTCAAGAGCACTGTGCAACA	132175		
RESULT	5				
AX354812		493 bp	DNA	linear	PAT 06-PEB-2002
LOCUS	AX354812				
DEFINITION	Sequence 6 from Patent WO0179290.				
ACCESSION	AX354812				
VERSION	AX354812.1	GI:18619543			
KEYWORDS	.				
SOURCE	Mus sp.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (sites)				
AUTHORS	Drucker,D.J. and Lovshin,J.A.				
TITLE	GIP-2 receptor gene promoter and uses thereof				
JOURNAL	Patent: WO 0179290-A 6 25-OCT-2001; 1149336 ONTARIO INC. (CA) Location/Qualifiers 1..493 /organism="Mus sp." /db_xref="taxon:10095" 401..>466 /note="unnamed protein product" /codon_start=1 /protein_id="CAD22892.1" /db_xref="GI:18619544"				

BASE COUNT	100 a	132 c	155 g	104 t	2 others
Query Match	10.5%	Score 175;	DB 6;	Length 493;	
Best Local Similarity	99.5%	Pred. No. 5,76-42;			
Matches 186;	Conservative	0;	Mismatches	0;	Indels 1; Gaps 1;
<p> QY 1474 ATGCTTCCTTTTCTTCTTGCGCTTGAGAGACTCCAGGACGCTGATGAGACTT 1533 Db 1 ATGCTTCCTTTTCTTCTTGCGCTTGAGAGACTCCAGGACGCTGATGAGACTT 60 QY 1534 GGTAGCTGTGGGAAAATCTCCCAAGATTTTATAGAGGGGCGAGCGGGGATGAGAACTT 1593 Db 61 GGTAGCTGTGGGAAAATCTCCCAAGA-TTTATAGAGGGGCGAGCGGGGATGAGAACTT 119 QY 1594 GGAGATTCGTAGATCAGCTGTAGAGCACTCAGACAGTCGGCGGCTGAAGAGACTTGT 1653 Db 120 GGAGATTCGTAGATCAGCTGTAGAGCACTCAGACAGTCGGCGGCTGAAGAGACTTGT 179 QY 1654 GCAACA 1660 Db 180 GCAACA 186 </p>					
RESULT 6					
AC026591/c	198271 bp	DNA	linear	HTG 27 AUG-2001	
LOCUS	Homo sapiens chromosome 17 clone RP11-655D3 map 17,			*** SEQUENCING	
DEFINITION	IN PROGRESS ***, 4 ordered pieces.				
ACCESSION	AC026591				
VERSION	AC026591.9 GI:15291083				
KEYWORDS	HTG; HTGS_PHASE2; HTGS_FUZZTOP; HTGS_ACTIVEPIN.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Birren, B., Linton, L., Nussbaum, C. and Lander, E.				
TITLE	1 (bases 1 to 198271)				
JOURNAL	Homo sapiens chromosome 17, clone RP11-655D3				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 198271)				
TITLE	Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, Y., Bede, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McSheeters, R., Melnick, J., Meneses, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanil, C., Pollari, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange, Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Testave, S., Theodore, J., Tirelli, A., Travers, M., Triggillo, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Aug 26, 2001 this sequence version replaced g1:14547870. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
	Center: Whitehead Institute/ MIT Center for Genome Research				
	Center code: MIBR				

Db 144074 CAAACA 144069

RESULT 7	AC087646/c	LOCUS	DEFINITION	AC087646	210719 bp	DNA	linear	HTG 31-MAY-2001
			Human sapiens chromosome 17 clone RP11-773021 map 17, WORKING DRAFT SEQUENCE, 12 unordered pieces.	AC087646				
				AC087646				
				AC087646.3	GI:14269760			
				HTG: HTGS_PHASE1, HTGS_DRAFT; HTGS_FULLTOP.				
				human.				
				Human sapiens				
				ORANISM				

Submitted (5-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 31, 2001 this sequence version replaced gi:13493095.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L11995
Center clone name: 773_O_21
----- Summary Statistics -----
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 205962 bases at least Q40
Consensus quality: 208431 bases at least Q30
Consensus quality: 209109 bases at least Q20
Insert size: 205000; agarose-fp
Insert size: 209619; sum-of-ctrls
Quality coverage: 8.4 in Q20 bases; agarose-fp
Quality coverage: 8.2 in Q20 bases; sum-of-ctrls

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as

Consensus quality: 203603 bases at least Q40	
Consensus quality: 206822 bases at least Q30	
Consensus quality: 208592 bases at least Q20	
Insert size: 214000; agarose-fp	
Insert size: 211859; sum-of-contigs	
Quality coverage: 5.34 in Q20 bases; agarose-fp	
Quality coverage: 5.44 in Q20 bases; sum-of-contigs	

* NOTE: This is a 'working draft' sequence. It currently	
* consists of 16 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	

1	1400: contig of 1400 bp in length
*	1401
*	1500: gap of unknown length
*	1501
*	4621: contig of 3121 bp in length
*	4622
*	4721: gap of unknown length
*	4722
*	7894: contig of 3173 bp in length
*	7895
*	7995
*	10806: contig of 2812 bp in length
*	10807
*	10906: gap of unknown length
*	16818: contig of 5912 bp in length
*	16819
*	16918: gap of unknown length
*	16919
*	24662: contig of 7744 bp in length
*	24663
*	24763
*	32389: contig of 7627 bp in length
*	32390
*	32489: gap of unknown length
*	40414: contig of 7925 bp in length
*	40415
*	40514: gap of unknown length
*	40515
*	53087: contig of 12573 bp in length
*	53088
*	53187: gap of unknown length
*	53188
*	64676: contig of 11489 bp in length
*	64677
*	64777
*	78336: contig of 13760 bp in length
*	78337
*	78636: gap of unknown length
*	78637
*	94647: contig of 16011 bp in length
*	94648
*	94748
*	117462: contig of 22715 bp in length
*	117463
*	117562: gap of unknown length
*	117563
*	143519: contig of 22957 bp in length
*	143520
*	143619: gap of unknown length
*	143620
*	171176: contig of 27557 bp in length
*	171177
*	171276: gap of unknown length
*	171277
*	213359: contig of 42083 bp in length.

FEATURES	
source	
1..213359	
Location/Qualifiers	
1..213359	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="RPC1-11"	
/clone="RP11-773021"	
1..1400	
misc_feature	
1501..4621	
/note="assembly_name:Contig18"	
misc_feature	
4722..7894	
/note="assembly_name:Contig19"	
misc_feature	
7995..10806	
/note="assembly_name:Contig20"	
misc_feature	
10907..16818	
/note="assembly_name:Contig21"	
misc_feature	
16913..24662	
/note="assembly_name:Contig22"	
misc_feature	
clone_end:17	
vector_side:right	
24763..32389	
/note="assembly_name:Contig24"	
misc_feature	
32490..40414	
/note="assembly_name:Contig25"	
misc_feature	
40515..53087	
/note="assembly_name:Contig26"	

	misc_feature	53188..64676	/note="assembly_name:Contig27"	
	misc_feature	64777..78536	/note="assembly_name:Contig28"	
	misc_feature	78637..94647	/note="assembly_name:Contig29"	
	misc_feature	94748..117462	/note="assembly_name:Contig30"	
	misc_feature	117563..143519	/note="assembly_name:Contig31"	
	misc_feature	143620..171116	/note="assembly_name:Contig32"	
	misc_feature	171277..213359	/note="assembly_name:Contig33"	
	misc_feature	213360..248599	/note="assembly_name:Contig34"	
	misc_feature	248600..284116	/note="assembly_name:Contig35"	
	misc_feature	284117..320633	/note="assembly_name:Contig36"	
	misc_feature	320634..357150	/note="assembly_name:Contig37"	
	misc_feature	357151..393667	/note="assembly_name:Contig38"	
	misc_feature	393668..430184	/note="assembly_name:Contig39"	
	misc_feature	430185..466701	/note="assembly_name:Contig40"	
	misc_feature	466702..503218	/note="assembly_name:Contig41"	
	misc_feature	503219..539735	/note="assembly_name:Contig42"	
	misc_feature	539736..576252	/note="assembly_name:Contig43"	
	misc_feature	576253..612769	/note="assembly_name:Contig44"	
	misc_feature	612770..649286	/note="assembly_name:Contig45"	
	misc_feature	649287..685803	/note="assembly_name:Contig46"	
	misc_feature	685804..722320	/note="assembly_name:Contig47"	
	misc_feature	722321..758837	/note="assembly_name:Contig48"	
	misc_feature	758838..795354	/note="assembly_name:Contig49"	
	misc_feature	795355..831871	/note="assembly_name:Contig50"	
	misc_feature	831872..868388	/note="assembly_name:Contig51"	
	misc_feature	868389..904905	/note="assembly_name:Contig52"	
	misc_feature	904906..941422	/note="assembly_name:Contig53"	
	misc_feature	941423..977939	/note="assembly_name:Contig54"	
	misc_feature	977940..1014456	/note="assembly_name:Contig55"	
	misc_feature	1014457..1050973	/note="assembly_name:Contig56"	
	misc_feature	1050974..1087490	/note="assembly_name:Contig57"	
	misc_feature	1087491..1124007	/note="assembly_name:Contig58"	
	misc_feature	1124008..1160524	/note="assembly_name:Contig59"	
	misc_feature	1160525..1197041	/note="assembly_name:Contig60"	
	misc_feature	1197042..1233558	/note="assembly_name:Contig61"	
	misc_feature	1233559..1270075	/note="assembly_name:Contig62"	
	misc_feature	1270076..1306592	/note="assembly_name:Contig63"	
	misc_feature	1306593..1343109	/note="assembly_name:Contig64"	
	misc_feature	1343110..1379626	/note="assembly_name:Contig65"	
	misc_feature	1379627..1416143	/note="assembly_name:Contig66"	
	misc_feature	1416144..1452660	/note="assembly_name:Contig67"	
	misc_feature	1452661..1489177	/note="assembly_name:Contig68"	
	misc_feature	1489178..1525694	/note="assembly_name:Contig69"	
	misc_feature	1525695..1562211	/note="assembly_name:Contig70"	
	misc_feature	1562212..1598728	/note="assembly_name:Contig71"	
	misc_feature	1598729..1635245	/note="assembly_name:Contig72"	
	misc_feature	1635246..1671762	/note="assembly_name:Contig73"	
	misc_feature	1671763..1708279	/note="assembly_name:Contig74"	
	misc_feature	1708280..1744796	/note="assembly_name:Contig75"	
	misc_feature	1744797..1781313	/note="assembly_name:Contig76"	
	misc_feature	1781314..1817830	/note="assembly_name:Contig77"	
	misc_feature	1817831..1854347	/note="assembly_name:Contig78"	
	misc_feature	1854348..1890864	/note="assembly_name:Contig79"	
	misc_feature	1890865..1927381	/note="assembly_name:Contig80"	
	misc_feature	1927382..1963898	/note="assembly_name:Contig81"	
	misc_feature	1963899..2000415	/note="assembly_name:Contig82"	
	misc_feature	2000416..2036932	/note="assembly_name:Contig83"	
	misc_feature	2036933..2073449	/note="assembly_name:Contig84"	
	misc_feature	2073450..2109966	/note="assembly_name:Contig85"	
	misc_feature	2109967..2146483	/note="assembly_name:Contig86"	
	misc_feature	2146484..2182999	/note="assembly_name:Contig87"	
	misc_feature	2183000..2219516	/note="assembly_name:Contig88"	
	misc_feature	2219517..2256033	/note="assembly_name:Contig89"	
	misc_feature	2256034..2292550	/note="assembly_name:Contig90"	
	misc_feature	2292551..2329067	/note="assembly_name:Contig91"	
	misc_feature	2329068..2365584	/note="assembly_name:Contig92"	
	misc_feature	2365585..2402101	/note="assembly_name:Contig93"	
	misc_feature	2402102..2438618	/note="assembly_name:Contig94"	
	misc_feature	2438619..2475135	/note="assembly_name:Contig95"	
	misc_feature	2475136..2511652	/note="assembly_name:Contig96"	
	misc_feature	2511653..2548169	/note="assembly_name:Contig97"	
	misc_feature	2548170..2584686	/note="assembly_name:Contig98"	
	misc_feature	2584687..2621203	/note="assembly_name:Contig99"	
	misc_feature	2621204..2657720	/note="assembly_name:Contig100"	
	misc_feature	2657721..2694237	/note="assembly_name:Contig101"	
	misc_feature	2694238..2730754	/note="assembly_name:Contig102"	
	misc_feature	2730755..2767271	/note="assembly_name:Contig103"	
	misc_feature	2767272..2803788	/note="assembly_name:Contig104"	
	misc_feature	2803		

Db 84362 CTTCCCAAGTACCTGCTGCTCAAGCAACTGTGTGAGAAAGACCTGCTATCCGGTTCTGCT 84303
OY 880 TTTTGTGCTGCTGACGCTTGACAAACACCCCTGCTTTGAACCCCACTATTTCTC 939
Db 84302 GTTATTGCTGCTGCTGCTTTGGGCAACCCATCCAACTCTCTGAGCTTCTCTC 84243
OY 940 AGCCCTCAGATGAAAGTAATGTA--CCTTGAGAGTAAGTATGATGGGTTCAAGTAAGTA 998
Db 84242 AGCTGCAAAATGAGAAATTCATGATCCTCATGGGTTATGTCAGAGATGTCGAAATA 84183
OY 999 GGGCAGAGGCTGGAAGTTTGTAAACCTAAACATGAAGTGGGCTTTGTTAGTAAGT 1058
Db 84182 GGATGGGCTGTAAGACG--TGTCATATGTAAGTTGTGGGACAGCCCTACTCTGAT 84127
OY 1059 CCATGATTAACCATTAATAATCTGTCAGAGTGCCAGACATCACTGTTGTGAACACAA 1118
Db 84126 GCTGTGATGTAAGTATGATATCATCAATTCATTAAGAGGTGTGACAGTGGCTGAGCC 84067
OY 1119 CGGCCACCTGAAACAGCGGCAATTAAGGACCAACCTCCGCTGCTTCTTACCCGTT 1178
Db 84066 ACATGCTTGAAGCAAGTGGGCAATTAAGATGATGATGTCATTTTCAACACACT 84007
OY 1179 GTTTTGTACCGTTAAAGCAACTCACTCCGCACTGAAAGAGCTTTTCTGACAGC 1238
Db 84006 GTTTTGTATGTTGTAGCATTAACCTA--CGCCGCTGATGGGTTTGTCTTCTCATGG 83952
OY 1239 CTGGGCTCTGAGGCTGTTGCTCTGAGACAGAAACATCACTTGTACTATGAGTACT 1298
Db 83951 CTGGGCTGAGGAGCTGTTCTCTGAGACTGAGCAACCACTCTCTGAGCTCTGCAAT 83892
OY 1299 AGTAACACGAGAGCTTGAGATAGCCCT-----GAGCTGGTCCCTTTAGAAAGTT 1350
Db 83891 AAACAGAGACCCAGAGAGATTTGCTTTGTGTTAGCTCAGCCCTGAGCAACGCT 83832
OY 1351 TGATGCTTGAAGAAATGCTGGCTTAAAGAG-----CCTACTGGCATAGGGG 1401
Db 83831 TAATGCTTGAAGAAATATACCTGCTCTCAAAAGACACCTGCTTAACCTGCTCGAG 83772
OY 1402 GCCCATCTCTCCAGCATCCGAATCTC-----AAT 1432
Db 83771 CCCCATGCGCTTTCGCCAGCATTCCTCTGTAGCTCGAAACCCCACTGCTCAGCCAC 83712
OY 1433 CTGCTGCTGCTGCTGAAGATGAAATCTCGGAATGCTGTAACCATGCTGCTTTCTTCT 1492
Db 83711 TTCTGCTGATGAGACGAACCTTCAAGGCTGAGACCGCTTGTCTTCTTCTCT 83652
OY 1493 GGGCTTCTGCTGAGAGTCCCGAGGAGCGTGAAGCTGTTGGGCTAGTCTGGAAAAATC 1552
Db 83651 CAGCTCTTCAAGGAAGTCCCGAAAGACACAGCTGACT--AGGGAAGTCTGGAAAAATC 83593
OY 1553 TCCCAAGATTTTGAAGGAGCGGAGGAGTGA-----GAAACTTGGAG 1597
Db 83592 TCCTGCTGTTTGGGGGAGGAGGAGGAGGAGTGAAGCCAGGCGGAGAAAGTCTGAA 83533
OY 1598 ATTGCTGATGCTGCTGTAAGACCAACTCAGACATC-----GGCGGCTTGAAGAGGACTTGT 1653
Db 83532 ACTCCCTGATGCTGCTTGAAGACCGCTCAGACACTCTCGGCGAGCGTGAAGAGATTTGT 83473
OY 1654 GCAACAA 1660
Db 83472 GCAACAA 83466

RESULT 9
LOCUS AC100859 49665 bp DNA linear HTG 22-NOV-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-615H7 map 15, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC100859
VERSION AC100859.1 GI:17048229
KEYWORDS HTG; HTGS_PHASED.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 49665)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-615H7
Unpublished
2 (bases 1 to 49665)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barron,A., Bastien,V., Boguslavsky,L., Boukhaltier,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choedel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,J., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Jillev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labrecque,K.,
Lamzares,R., Landers,T., Lehocck,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21610
Center clone name: 615_H_7

* NOTE: This record contains 61 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
686 785: contig of 685 bp in length
786 1524: gap of 100 bp
1525 1624: contig of 739 bp in length
1625 2300: contig of 676 bp in length
2301 2400: gap of 100 bp
2401 3073: contig of 673 bp in length
3074 3173: gap of 100 bp
3174 3870: contig of 697 bp in length
3871 3970: gap of 100 bp
3971 4667: contig of 697 bp in length
4668 4767: gap of 100 bp
4768 5508: contig of 741 bp in length
5509 5608: gap of 100 bp
5609 6347: contig of 739 bp in length
6348 6447: gap of 100 bp
6448 7196: contig of 749 bp in length
7197 7296: gap of 100 bp

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

HTG: HTGS-PHASE1.
Norway rat.
Rattus norvegicus
Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae;
Rattus

1 (bases 1 to 102783)

Mozny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alstbooks,S.L., Anaratunga,H.C., Are,J.R., Banks,T., Barbata,J.,
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burdett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy,Carroll,L., Dederich,D.A., Delaney,K.R., Delaper,O.,
Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Hollway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewins,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Matlhez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,
Ogih,M., Okunou,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoshitai,N.,
Slason,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellro,B., Thomas,N.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Wellington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 102783)
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16901736.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GEMB

Center clone name: CH230-11B19

Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 92140 bases at least Q40
Consensus quality: 98027 bases at least Q30
Consensus quality: 102801 bases at least Q20
Estimated insert size: 79332; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 6683: contig of 6683 bp in length
6684 6783: gap of unknown length
6784 12614: contig of 5831 bp in length
12615 12714: gap of unknown length
12715 16628: contig of 3914 bp in length
16629 16729: gap of unknown length
16729 19333: contig of 2604 bp in length
19333 19433: gap of unknown length
19433 23463: contig of 4030 bp in length
23463 23562: gap of unknown length
23562 25954: contig of 2392 bp in length
25954 26054: gap of unknown length
26054 27900: contig of 1845 bp in length
27900 27999: gap of unknown length
27999 30376: contig of 2377 bp in length
30376 30476: gap of unknown length
30476 33824: contig of 3348 bp in length
33824 33924: gap of unknown length
33924 36244: contig of 2220 bp in length
36244 36444: gap of unknown length
36444 37543: contig of 1299 bp in length
37543 37644: gap of unknown length
37644 39819: contig of 2176 bp in length
39819 39919: gap of unknown length
39919 41518: contig of 1599 bp in length
41518 41618: gap of unknown length
41618 43294: contig of 1676 bp in length
43294 43394: gap of unknown length
43394 45183: contig of 1789 bp in length
45183 45283: gap of unknown length
45283 46781: contig of 1498 bp in length
46781 46881: gap of unknown length
46881 48828: contig of 1947 bp in length
48828 48928: gap of unknown length
48928 50954: contig of 2026 bp in length
50954 51054: gap of unknown length
51054 52624: contig of 1570 bp in length
52624 52724: gap of unknown length
52724 54323: contig of 1599 bp in length
54323 54423: gap of unknown length
54423 55973: contig of 1550 bp in length
55973 56073: gap of unknown length
56073 57987: contig of 1914 bp in length
57987 58087: gap of unknown length
58087 60090: contig of 2003 bp in length
60090 60190: gap of unknown length
60190 61729: contig of 1539 bp in length
61729 61829: gap of unknown length
61829 62932: contig of 1103 bp in length
62932 63032: gap of unknown length
63032 64638: contig of 1606 bp in length
64638 64738: gap of unknown length
64738 66604: contig of 1866 bp in length
66604 66704: gap of unknown length
66704 68585: contig of 1881 bp in length
68585 68685: gap of unknown length
68685 70000: contig of 1315 bp in length
70000 72122: gap of unknown length
72122 72222: contig of 2022 bp in length
72222 74384: contig of 2162 bp in length
74384 74484: gap of unknown length
74484 75864: contig of 1380 bp in length

*	75865	75964:	gap of unknown length
*	75865	77361:	contig of 1397 bp in length
*	77362	77461:	gap of unknown length
*	77462	79807:	contig of 2346 bp in length
*	79808	79907:	gap of unknown length
*	79908	81416:	contig of 1509 bp in length
*	81417	81516:	gap of unknown length
*	81517	82905:	contig of 1389 bp in length
*	82906	83005:	gap of unknown length
*	83006	84278:	contig of 1273 bp in length
*	84279	84378:	gap of unknown length
*	84379	85753:	contig of 1375 bp in length
*	85754	85853:	gap of unknown length
*	85854	87704:	contig of 1851 bp in length
*	87705	87804:	gap of unknown length
*	87805	88814:	contig of 1010 bp in length
*	88815	88914:	gap of unknown length
*	88915	90378:	contig of 1464 bp in length
*	90379	90379:	gap of unknown length
*	90479	91537:	contig of 1059 bp in length
*	91538	91637:	gap of unknown length
*	91638	92800:	contig of 1163 bp in length
*	92801	92900:	gap of unknown length
*	92901	94274:	contig of 1374 bp in length
*	94275	94374:	gap of unknown length
*	94375	95621:	contig of 1247 bp in length
*	95622	95721:	gap of unknown length
*	95722	96844:	contig of 1123 bp in length
*	96845	96944:	gap of unknown length
*	96945	98084:	contig of 1140 bp in length
*	98085	98184:	gap of unknown length
*	98185	99274:	contig of 1090 bp in length
*	99275	99374:	gap of unknown length
*	99375	100542:	contig of 1168 bp in length
*	100543	100642:	gap of unknown length
*	100643	101651:	contig of 1009 bp in length
*	101652	101751:	gap of unknown length
*	101752	102783:	contig of 1032 bp in length

[illegible]

Query Match	4.4%;	Score 73.2;	DB 2;	Length 102783;
Best Local Similarity	66.5%;	Pred. No. 7.5e-11;		
Matches 105;	Conservative	0;	Mismatches 53;	Indels 0;
				Gaps 0;

Oy	225	CAAAACACACACACACACACACACACACACACACACATGTTT	284
Dd	72488	CAGCACACACACACACACACACACACACACACACACAGCTGC	72547
Oy	285	TCTTCCTCCCTCCACTTCCTCCCATTCCTGTGGGCCCAAGAGATGCATTATGACT	344
Dd	72548	ACATGCTCTTCACAAAGCATCATGCGCCTCTAATGGAGGCCACAGCTGTCTTCATGTGGACA	72607
Oy	345	GTAGAAATCACACCACCATTAAGGCCCATCTGGAGGCC	382
Dd	72608	TTAAGGAACAGGAATTAATAGAAGACTTGTCGCCAGCC	72645

RESULT 11	AC095179/c	171332 bp	DNA	linear	HNG 20-DEC-2001
LOCUS	AC095179	171332 bp	DNA	linear	HNG 20-DEC-2001
DEFINITION	Rattus norvegicus clone CH230-9C13, 64 unordered pieces.				
ACCESSION	AC095179				
VERSION	AC095179.2	GI:17942259			
KEYWORDS	HTG; HTGS; PHASE1.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				

Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 171322)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alstrooks,S.L., Amarantunga,H.C., Are,J.R., Banks,T., Barbarella,J., Benton,J., Blevins,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C., Bowie,S., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chaver,D., Chen,G., Chen,R., Chen,Z., Chowhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhajj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunnarane,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,K.R., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulys,K.S., Hume,J., Johnson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratovic,J., Kreshl,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,Y., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawlinney,E., Mcloed,M.P., Meador,M., Meil,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickelson,E., Nwokenko,S., Ogun,M., Okumotu,G., Oraqunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Peteru,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubokan,I., Rolfe,M., Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoochari,N., Stinson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmali,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williams,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 171322)

Worley,K.C.

Direct Submission

Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15625733.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GDAG

Center clone name: CH230-9C13

Summary Statistics

Assembly program: Phrap; version 0.990329First call to findhaplist

Consensus quality: 147096 bases at least Q40

Consensus quality: 154402 bases at least Q30

Consensus quality: 160566 bases at least Q20

Estimated insert size: 145980; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-ip estimation

Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

NOTE: This is a 'working draft' sequence. It currently

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 72930)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-116F12
Unpublished
2 (bases 1 to 72930)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukharov, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choe, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPeckers, R., Meldrum, J.,
Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainou, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L16561
Center clone name: 116_F_12

* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
729 828: contig of 728 bp in length
829 828: gap of 100 bp
1547 1646: contig of 718 bp in length
1647 1647: gap of 100 bp
2375 2474: contig of 728 bp in length
2475 3195: contig of 721 bp in length
3196 3295: gap of 100 bp
3296 4033: contig of 738 bp in length
4034 4133: gap of 100 bp
4134 4866: contig of 733 bp in length
4867 4966: gap of 100 bp
4967 5694: contig of 728 bp in length
5695 5794: gap of 100 bp
5795 6520: contig of 726 bp in length
6521 6620: gap of 100 bp
6621 7350: contig of 730 bp in length
7351 7450: gap of 100 bp

7451 8152: contig of 702 bp in length
8153 8252: gap of 100 bp
8253 8968: contig of 716 bp in length
8969 9068: gap of 100 bp
9069 9775: contig of 707 bp in length
9776 9875: gap of 100 bp
9876 10583: contig of 708 bp in length
10584 10683: gap of 100 bp
10684 11397: contig of 714 bp in length
11398 11497: gap of 100 bp
11498 12214: contig of 717 bp in length
12215 12314: gap of 100 bp
12315 13043: contig of 729 bp in length
13044 13143: gap of 100 bp
13144 13881: contig of 728 bp in length
13882 13981: gap of 100 bp
13982 14717: contig of 736 bp in length
14718 14817: gap of 100 bp
14818 15552: contig of 735 bp in length
15553 15652: gap of 100 bp
15653 16382: contig of 730 bp in length
16383 16482: gap of 100 bp
16483 17201: contig of 719 bp in length
17202 17301: gap of 100 bp
17302 18013: contig of 712 bp in length
18014 18113: gap of 100 bp
18114 18825: contig of 713 bp in length
18827 18926: gap of 100 bp
18927 19648: contig of 722 bp in length
19649 19748: gap of 100 bp
19749 20470: contig of 722 bp in length
20471 20570: gap of 100 bp
20571 21302: contig of 732 bp in length
21303 21402: gap of 100 bp
21403 22113: contig of 711 bp in length
22114 22213: gap of 100 bp
22214 22937: contig of 724 bp in length
22938 23037: gap of 100 bp
23038 23752: contig of 715 bp in length
23753 23852: gap of 100 bp
23853 24571: contig of 719 bp in length
24572 24671: gap of 100 bp
24672 25395: contig of 724 bp in length
25396 25495: gap of 100 bp
25496 26207: contig of 712 bp in length
26208 26307: gap of 100 bp
26308 27035: contig of 728 bp in length
27036 27135: gap of 100 bp
27136 27854: contig of 719 bp in length
27855 27954: gap of 100 bp
27955 28664: contig of 710 bp in length
28665 28764: gap of 100 bp
28765 29468: contig of 704 bp in length
29469 29568: gap of 100 bp
29569 30300: contig of 732 bp in length
30301 30400: gap of 100 bp
30401 31134: contig of 734 bp in length
31135 31234: gap of 100 bp
31235 31970: contig of 736 bp in length
31971 32070: gap of 100 bp
32071 32798: contig of 728 bp in length
32799 32898: gap of 100 bp
32899 33517: contig of 619 bp in length
33518 33617: gap of 100 bp
33618 34307: contig of 690 bp in length
34308 34407: gap of 100 bp
34408 35131: contig of 724 bp in length
35132 35231: gap of 100 bp
35232 35958: contig of 727 bp in length
35959 36058: gap of 100 bp
36059 36796: contig of 738 bp in length
36797 36896: gap of 100 bp
36897 37600: contig of 704 bp in length

[illegible]

	Query Match	4.3%	Score	72.2;	DB 2:	Length	72930;		
	Best Local Similarity	40.28;	Pred. No.	1.5e-10;					
	Matches	107;	Conservative	0;	Mismatches	159;	Indels	0; Gaps	0
Oy	138	AATATTGTGACAAACAGCGTCTTTTCGCCCCACCATGGAGCACC6GAGCTCCACTGATA	197						
Db	47748	AATGAGTCACTACGAATCCAGTGAAATCCCCATCAAATATCCACGCCAATTTTTGCAGAAC	47689						
Oy	198	TGGACAGAATAGCTTTACAGTCACTTCAAAAACACACACACACACACACACACACA	257						
Db	47688	AAGAAAGCGTAATTTCCAACCTCATATGSAAMAAACACACACACACACACACACACACA	47629						
Oy	258	CACACACACACACACACACACATGTTTTCTTCCCTCCCTCGACTCTCCCATTCCTCT	317						
Db	47628	CACACACACACACACACACAGGAATTTCCACACANNNNNNNNNNNNNNNNNNNNN	47569						
Oy	318	GGTCCCAAGAGATGACCATATTGACTGTAGAAATCACACACATATAAACCATCTGG	377						
Db	47568	NN	47509						
Oy	378	GAGCATTTCCAGACTGATCTTTTAA	403						
Db	47508	NNNNNNNNNNNNNNNNNNNNAAAAATTTA	47483						

RESULT 13	AF417202	17529 bp	DNA	linear	ROD 11-OCT-2001
LOCUS	AF417202	17529 bp	DNA	linear	ROD 11-OCT-2001
DEFINITION	Mus musculus strain 129/SvJ chromosome 6 map 6f1-f2 sequence.				
ACCESSION	AF417202				
VERSION	AF417202.1	GI:16033585			
KEYWORDS					
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 17529)				
TITLE	Lee, C.-M. and Chang, Y.-S.				
JOURNAL	129/SvJ mouse genomic clone				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 17529)				
TITLE	Lee, C.-M. and Chang, Y.-S.				
JOURNAL	Direct Submission				
FEATURES	Submitted (06-SEP-2001) Graduate Institute of Basic Medical Sciences, Chang-Gun University School of Medicine, 259 Wen-Hua 1st Road, Kwei-Shan, Tao-Yuan, Taiwan 033, Republic of China				
source	1..17529				
BASE COUNT	5160 a 3937 c 3800 g 4632 t				
ORIGIN					
Query Match	4.38; Score 71.2; DB 10; Length 17529;				
Best Local Similarity	67.6%; Ped. No. 2, 6e-10;				
Matches 100; Conservative	0; Mismatches 48; Indels 0; Gaps 0;				
Y 227	AAACGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACATGTTTC 286				
DB 8770	AAACGACATACGACAGACAGACAGACAGACAGACAGACAGACAGACATATAT 8829				
Y 287	TTCCCTCCCTCCGACTTCTCTCCCAFTTCTGTGGTCCCAAGATGACATATTGACT 346				
DB 8830	ATATCTGAAACCAACCACTATGAATCTCTAGTGGGAAATAAAGAGCTGAGACTAAAGCCGG 8889				
Y 347	AGAAATGACACGACCATTAAGCCCATC 374				
DB 8890	TGTGATGTACAGCGCTATATCCGACG 8917				
RESULT 14					
AC099092	176047 bp	DNA	linear	HTG 21-DEC-2001	
LOCUS	AC099092				
DEFINITION	Rattus norvegicus clone CH30-109013, *** SEQUENCING IN PROGRESS				
ACCESSION	AC099092				
VERSION	AC099092.2	GI:17975725			
KEYWORDS	HTG; HTGS; PHASEL.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
AUTHORS	1 (bases 1 to 176047)				
LOCUS	Murny, D.M., Adams, C., Adlo-Oduola, B., Ali-oshan, F.R., Allen, C., Alibrooks, S.L., Amerasinghe, H.C., Are, J.R., Banks, T., Barbarella, J., Benton, J., Blinze, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Blevins, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, L., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathore, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Demu, A.L., Ding, Y., Dinu, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,				

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Gatt,R., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtharge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabab,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenwo,S., Ogutu,N., Okunoye,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primm,E., Pu,L.L., Ruitz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,H., Sutton,A., Sytek,A., Tabor,P., Tamerisa,A., Stanley,H., Thomas,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,K., Tang,H., Tansey,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Walli,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Unpublished

2 (bases 1 to 176047)

Worley,K.C.

Direct Submission

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:16874566.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: G6NG

Center clone name: CH230-189013

Summary Statistics

Assembly program: Phrap; version 0.990329Pfirst call to findPhrapList

Consensus quality: 151055 bases at least Q40

Consensus quality: 158732 bases at least Q30

Consensus quality: 165129 bases at least Q20

Estimated insert size: 160905; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.5x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see <http://www.hgsc.bcm.tmc.edu/docs/genbank.draft.data.html>).

NOTE: This is a 'working draft' sequence. It currently
consists of 42 contigs. The true order of the pieces

is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence

as soon as it is available and the accession number will
be preserved.

1 12798: contig of 12798 bp in length
12898: gap of 12898 bp in length
12899: gap of 12899 bp in length
25237: contig of 12329 bp in length
25238: gap of unknown length
25328: gap of unknown length
37905: contig of 12578 bp in length
38005: gap of unknown length
37906: gap of unknown length
45387: contig of 7382 bp in length
45388: gap of unknown length

45488 52837: contig of 7350 bp in length
52838 52937: gap of unknown length
52938 60684: contig of 7747 bp in length
60685 60785: gap of unknown length
60785 67091: contig of 6307 bp in length
67091 73887: gap of unknown length
73887 73987: contig of 6696 bp in length
73987 82832: gap of unknown length
82832 82932: gap of 8844 bp in length
82932 87168: contig of 4237 bp in length
87168 92058: gap of unknown length
92058 92158: contig of 4790 bp in length
92158 97933: gap of unknown length
97933 97934: gap of 5775 bp in length
97934 101431: gap of unknown length
101431 101432: contig of 3398 bp in length
101432 101531: gap of unknown length
101531 105158: contig of 3627 bp in length
105158 105259: gap of unknown length
105259 10532: gap of 5124 bp in length
10532 110382: contig of unknown length
110382 110482: gap of unknown length
110482 115931: gap of 5449 bp in length
115931 116031: contig of unknown length
116031 119788: gap of unknown length
119788 119888: contig of 3757 bp in length
119888 123331: gap of unknown length
123331 123332: contig of 3443 bp in length
123332 123432: gap of unknown length
123432 126326: contig of 2895 bp in length
126326 126426: gap of unknown length
126426 128477: contig of 2050 bp in length
128477 128576: gap of unknown length
128576 131745: contig of 3169 bp in length
131745 131746: gap of unknown length
131746 131846: contig of 4266 bp in length
131846 136211: gap of unknown length
136211 136212: contig of 3326 bp in length
136212 139537: gap of unknown length
139537 141833: gap of unknown length
141833 141834: contig of 2196 bp in length
141834 144814: gap of unknown length
144814 144914: contig of 2881 bp in length
144914 148363: gap of unknown length
148363 148463: contig of 3449 bp in length
148463 150147: gap of unknown length
150147 150247: contig of 1684 bp in length
150247 152674: gap of unknown length
152674 152774: contig of 2427 bp in length
152774 154591: gap of unknown length
154591 154691: contig of 1817 bp in length
154691 156159: gap of unknown length
156159 156259: contig of 1468 bp in length
156259 157478: gap of unknown length
157478 157578: contig of 1219 bp in length
157578 158858: gap of unknown length
158858 158959: contig of 1280 bp in length
158959 161046: gap of unknown length
161046 161146: contig of 2088 bp in length
161146 163452: gap of unknown length
163452 163552: contig of 2306 bp in length
163552 166098: gap of unknown length
166098 166198: contig of 2546 bp in length
166198 167902: gap of unknown length
167902 168002: contig of 1704 bp in length
168002 169409: gap of unknown length
169409 169509: contig of 1407 bp in length
169509 170627: gap of unknown length
170627 170727: contig of 1118 bp in length
170727 172694: gap of unknown length
172694 172794: contig of 1967 bp in length
172794 173803: gap of unknown length
173803 173903: contig of 1009 bp in length
173903 174905: gap of unknown length
174905 174905: contig of 1002 bp in length

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

